Image Cover Sheet

Défense nationale National Defence

UNCLASSIFIED

DEFENCE RESEARCH ESTABLISHMENT
CENTRE DE RECHERCHES POUR LA DÉFENSE
VALCARTIER, QUÉBEC

DEFENSIO PER SCIENTIAM

RESEARCH AND DEVELOPMENT BRANCH
DEPARTMENT OF NATIONAL DEFENCE
CANADA
BUREAU - RECHERCHE ET DÉVELOPPEMENT
MINISTÈRE DE LA DÉFENSE NATIONALE

Canada

SANS CLASSIFICATION

UNCLASSIFIED

DEFENCE RESEARCH ESTABLISHMENT
CENTRE DE RECHERCHES POUR LA DÉFENSE
VALCARTIER, QUÉBEC

DREV - R - 9527

Unlimited Distribution / Distribution illimitée

FUSION OF HIERARCHICAL IDENTITY DECLARATIONS FOR NAVAL COMMAND AND CONTROL

by

L. Des Groseilliers, É. Bossé and J. Roy

September / septembre 1996

Approved by / approuvé par

Chief Scientist / Scientifique en chef

6/9/96

Date

SANS CLASSIFICATION

ABSTRACT

Within the context of naval warfare, commanders and their staffs require access to a wide range of information to carry out their duties. This information provides them with the knowledge necessary to determine the position, identity and behavior of the enemy. This document is concerned with the fusion of identity declarations through the use of statistical analysis rooted in the Dempster-Shafer theory of evidence. It proposes to hierarchically structure the declarations according to STANAG 4420 (Display Symbology and Colours for NATO Maritime Units). More specifically the aim of this document is twofold: to explore the problem of fusing identity declarations emanating from different sources, and to offer the decision maker a quantitative analysis based on statistical methodology that can enhance his/her decision making process regarding the identity of detected objects.

RÉSUMÉ

Dans un contexte de guerre navale, le commandant d'un navire s'appuie sur un éventail imposant d'information pour analyser la situation tactique. Cette information comporte notamment des renseignements concernant la position, l'identité et le comportement des objets détectés. Ce document se concentre sur la fusion de déclarations d'identité au moyen de la théorie de l'évidence de Demspter-Shafer. On y propose de structurer de façon hiérarchique les déclarations d'identité à l'aide de la norme STANAG 4420 de l'OTAN. Une étude de la problématique de la fusion de déclarations d'identité provenant de plusieurs sources nous amène à offrir au commandant une analyse quantitative basée sur une méthodologie statistique pouvant le/la seconder dans son processus de prise de décision quant à l'identité des objets détectés.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

In today's naval warfare, commanders and their staffs, who are both users and active elements of command and control systems, require access to a wide range of information to carry out their duties. In particular, their actions are based on information concerning the position, identity and behavior of other vessels in their vicinity. The position information determines where objects are, whereas the identity information determines what they are. Behavioral information is concerned with what the objects are doing. In warfare, no one piece of information can be accepted as complete truth. In order to lessen the damaging effects of poor quality evidence, the combination of information from every possible source is of primary importance. This combination process has often been carried out manually but in order to cope with the ever increasing flow of information, automation has surfaced as a possible option for the fusion of positional and identity information.

This document is concerned with the use of the Dempster-Shafer theory of evidence for the fusion of identity declarations within a naval environment. It proposes to hierarchically structure the identity declarations according to NATO's STANAG 4420 charts, which provide a better base for achieving interoperability in information exchange between nations than uncontrolled alternatives.

The Bayesian approach is also investigated but is found to suffer from major deficiencies in a hierarchical context, when fully specified likelihoods are not available. Other problems associated with this approach are the coding of ignorance and the strict requirements on the belief of a hypothesis and its negation.

One drawback of the Dempster-Shafer evidential theory is the long calculation time required by its high computational complexity. Due to the hierarchical nature of the evidence, an algorithm proposed by Shafer & Logan is implemented which reduces the calculations from exponential to linear time proportional to the number of nodes in the tree. A semi-automated decision making technique, based on belief and plausibility values, is then described to select alternatives which best support the combined identity declarations. The final decision will be taken by the decision maker, because he/she remains an important part of the process and because the choice of the final identity is typically scenario and mission dependent.

This document has only begun to investigate the use of the Dempster-Shafer approach in the naval environment. In fact, the various concepts studied could be applicable to the domain of wide area fusion within the framework of a Communications, Command, Control and Intelligence (C^3I) system.

1.0 INTRODUCTION

In today's naval warfare, commanders and their staffs, who are both users and active elements of command and control systems, require access to a wide range of information to carry out their duties. In particular, their actions are based on information concerning the position, identity and behavior of other vessels in their vicinity (Wilson, Ref. 1). The position information determines where objects are, whereas the identity information determines what they are. Behavioral information is concerned with what the objects are doing.

In warfare, no one piece of information can be accepted as complete truth. In order to lessen the damaging effects of poor quality evidence, the combination of information from every possible source is of primary importance. This combination process has often been carried out manually but, in order to cope with the ever increasing flow of information, automation has surfaced as a possible option for the fusion of positional and identity information.

This document is concerned with the investigation of automated identification techniques through the use of statistical analysis rooted in the Dempster-Shafer theory of evidence. More specifically, the aim of this document is twofold: to explore the problem of fusing identity information emanating from different sources, and to offer the decision maker a quantitative analysis based on statistical methodology that can enhance his/her decision making process regarding the identity of detected objects.

Chapter 2 describes the problem facing naval commanders and gives a brief survey of current identity information sources available on a Canadian Patrol Frigate type ship for above water warfare. Potential future identity information sources are also mentioned. Three levels of information fusion architectures are also discussed which correspond to the three following categories of identity information: sensor signals, attribute information and identity declaration. As focus is brought on the fusion of identity declarations, it is suggested that NATO's STANAG 4420 (STAndard NATO AGreement) charts, which were designed for representing maritime tactical information, would be an appropriate tool for a hierarchical structuring of identity declarations.

Fusion approaches are then suggested based on the premise that identity declarations are probabilistic in nature, so that each declaration is characterized by a confidence value.

Chapters 3 and 4 describe two approaches capable of fusing uncertain information: the Bayesian paradigm of probability theory and the Dempster-Shafer evidential theory, respectively. These approaches are described in terms of standard and hierarchically structured information, and examples are offered. Two techniques for combining hierarchical information are detailed: the first is due to Pearl (Ref. 2) and the second developed by Shafer & Logan (Ref. 3). Advantages and disadvantages of each approach are discussed.

Chapter 5 briefly discusses decision making techniques based on the Dempster-Shafer representation and pertaining to information structured in a hierarchical manner, in order to provide a decision making approach to the identity declaration fusion problem.

Chapter 6 proposes an identity declaration fusion function based on the findings of Chapters 2 to 5. A complete example is provided; it details the inputs, the fusion results and decision making alternatives.

The research and development activities described in this document were performed at DREV between 1993 and 1995 under PSC12C.

2.0 PROBLEM DESCRIPTION

In today's naval warfare, commanders and their staffs, who are both users and active elements of command and control systems, require access to a wide range of information to carry out their duties. In particular, their actions are based on information concerning the position, identity and behavior of other vessels in their vicinity (Wilson, Ref. 1).

The prime parameter is the position of objects surrounding a ship because identity and behavior mean little unless they can be associated with position. The position information determines where objects are, whereas the identity information determines what they are. The third type of information is behavior, that is, finding out what the objects are doing in order to assess the potential threat. Deductive reasoning plays a key role in determining behavioral information (Wilson, Ref. 1).

In warfare, no one piece of information can be accepted as complete truth. In order to lessen the damaging effects of poor quality evidence, the combination of information from all available sources is of primary importance. This combination process has often been carried out manually but, in order to cope with the ever increasing flow of information, automation has surfaced as a possible option. This is particularly true for the fusion of positional information, but the same approach could also be considered for identity information.

Waltz & Llinas (Ref. 4) state that identity estimation is a much broader problem than positional estimation because identity is a much broader concept than position, involving a larger number of variables. Thus to better understand the identity fusion problem, one needs to look at these variables in terms of the origins and types of identity information. The following section gives examples of sources producing identity information.

2.1 Information Sources

In a maritime environment, various surveillance systems, electronic intelligence and human observations are examples of information sources available to the commander. Two types of source can be distinguished: organic and non-organic sources. When the tactical picture is formed from data gathered by sources under the jurisdiction of the commander, these sources are called organic. However, additional information is sometimes supplied by sources outside the jurisdiction of the commander; these are referred to as non-organic sources (Gibson, Ref. 5). Output from these sources is partitioned according to the type of information they provide; output data may be characterized as either positional or identity information.

Positional information represents the dynamic parameters describing the movement associated with an object (contact). This generally includes position, velocity and acceleration. *Identity information* can be defined as declarations, propositions or statements that contribute to establish the identity of an object (Refs. 6-7). Equivalently, identity information may be seen as information from various sources that helps in distinguishing one object from another. Possible values for identity information can span the range from *sensor signals*, to *attributes*, to *identity declarations*, as depicted in Fig. 1. The sensor signals represent some characteristics of the energy sensed. Attributes such as size, shape, degree of symmetry, emitter type, etc. are inferred from these characteristics. Identity declarations specify the detected object; in the Canadian Navy, for example, they can consist of a general classification of which the observed object is a member (surface combatant), a specific type of ship (frigate), a specific class (City Class) or a unique identity (Ville de Québec). Therefore surface combatant, frigate, City Class and Ville de Québec are all examples of identity declarations. Identity declarations can also include information concerning the threat designation of an object: pending, unknown, assumed friend, suspect, friend, neutral or hostile. It is noteworthy that for some authors such as Filippidis & Schapel (Ref. 8), the term identity refers only to the threat designation of an object. Within the context of this study, these threat designations will be classified under "threat category," which is, as mentioned earlier, a subdivision of identity declarations.

2.1.1 Current Information Sources

Organic sources available on a Canadian Patrol Frigate (CPF) type ship for above water warfare include surveillance and tracking radars, Electronic Warfare systems, Identification Friend or Foe (IFF) systems, operator intervention and link data exchanged by radio links among a group of platforms under the same command.

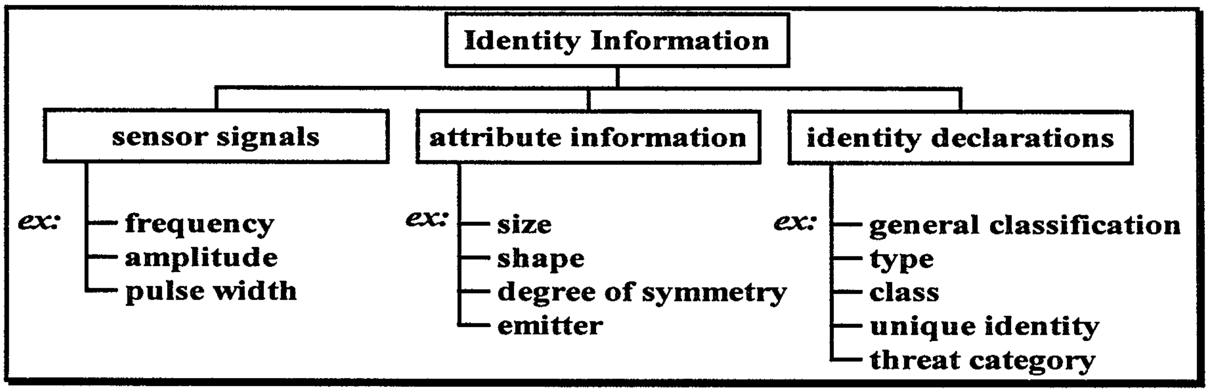


FIGURE 1 - Identity Information

In general, radars provide positional information in terms of range, azimuth and velocity components. Electronic Warfare (EW) systems include the Electronic Support Measure (ESM) and the Communication Intercept System (CIS). The ESM system intercepts electromagnetic radiation from active emitters in the environment and attempts to measure or estimate such parameters as angle of arrival, frequency, pulse repetition frequency, pulse width and scan period. These measurements are then compared to known characteristics of radar transmitters. The ESM system thus provides positional information (azimuth) as well as attribute information in the form of emitter type. It may also infer the platform identity (the object which contains the emitter) by matching the emitter type to a platform data base. The CIS system is capable of responding to programmable tasking to search the communication radio frequencies and provide bearings from airborne and surface originated radio frequency signals. The IFF system provides information (both in terms of position and identity) about a target when a cooperative target has responded to the interrogation. In the absence of an answer, only the location that delimits the sector in which the interrogation was performed is available but identity declarations may be inferred (Refs. 7, 9). Human intervention occurs when the operator of the CIS system listens to the signals and tries to estimate any attribute information and/or identity declarations (Ref. 7).

Intelligence reports and information from communication links are examples of non-organic sources from which positional and identity information can be obtained.

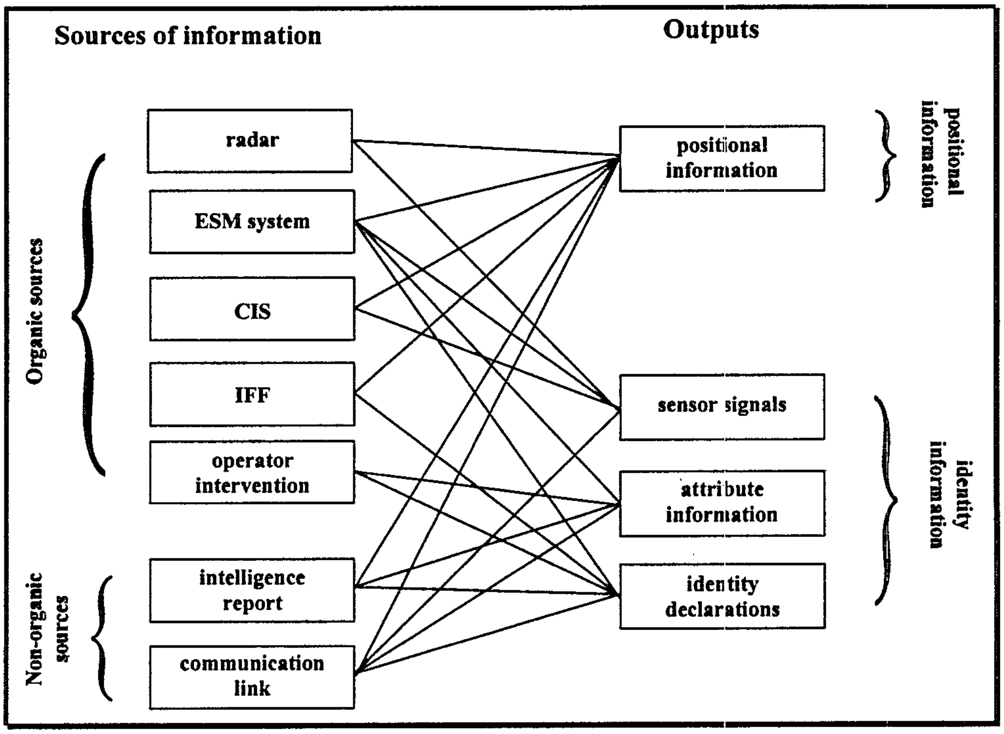


FIGURE 2 - Examples of Information Sources and Corresponding Outputs

Figure 2 gives examples of information sources with their corresponding outputs. The latter are defined in terms of positional information and identity information, which is itself subdivided into three categories: sensor signals, attribute information and identity declarations.

2.1.2 Advanced Information Sources

To cope with the increasing level of threat, sophisticated sensor processing is being developed to provide the commander with more accurate and timely information concerning the position and identity of detected objects. Examples of such technology advancement follow.

Some of the most promising approaches to object identification by active radars include: (1) inverse-synthetic aperture radar, which gives a two-dimensional image of an object; (2) radar-signal modulation, in which Doppler modulation of the signal provides target-specific information; (3) resonance response to short radar pulses; and (4) high-range-resolution radars (Ref. 10). It is expected that future radar systems will be capable of providing attribute information. Also, the position and speed calculated by radars could be used to infer attribute information.

Many countries are involved in the development of advanced naval EW systems (Refs. 11-12). The aim of these systems is to integrate all existing onboard electronic warfare systems - advanced electronic support measures (ESM), jammers, passive and active decoys - to provide a fully coordinated soft kill management. It is anticipated that the advanced ESM system will be able to identify the emitter type as well as the platform with a good confidence level.

Infrared imaging sensors afford another option for obtaining identity information. These devices sense the electromagnetic spectrum in the 3 to 12 μm waveband. They can automatically acquire small air objects; however, the classification and identification is mostly done manually. Studies are ongoing to automate these two processes (Ref. 13).

These advanced information sources can and will eventually produce attribute information and identity declarations in an autonomous fashion; in that sense, they are self- contained. Because of their sophisticated technology, they require comprehensive libraries of own and enemy signatures as well as powerful processing capabilities.

Figure 3 gives a more complete summary of information sources with their corresponding outputs. The bold lines indicate the differences between Figs. 2 and 3. A considerable amount of information pertaining to each category of identity information will be available. Figure 3 is not exhaustive in terms of information sources and position/identity information. Its aim is mainly to demonstrate the potentially high quantity of identity information that will eventually become available to estimate the identity of objects surrounding a ship.

2.2 Fusion of Identity Information

In a situation of intense activity, it would seem impossible to handle manually the flow of information pertaining to the position and identity of objects. Automation offers a viable solution. Hall (Ref. 6) states that fusion of identity information from multiple sources yields both qualitative and quantitative benefits because it takes advantage of the relative strengths of each source, resulting in an improved estimate of the object's identity. Quantitative benefits include increased confidence, e.g., higher probability of correct identification.

In order to automate the fusion of identity information, techniques must be devised to combine identity information at various levels: sensor signals, attribute information and identity declarations. This can be accomplished by focusing on the sensor processing units: energy, signal and target processing.

The energy processing unit is responsible for transforming the sensed energy into a signal, a form more suitable for target detection. Theoretically, the output of this stage can be sent to a fusion processor. This procedure is known as the *signal level* architecture, whereby signals from sensors are combined. An example of application is the fusion of pixels from imaging sensors to produce a single image (Ref. 14).

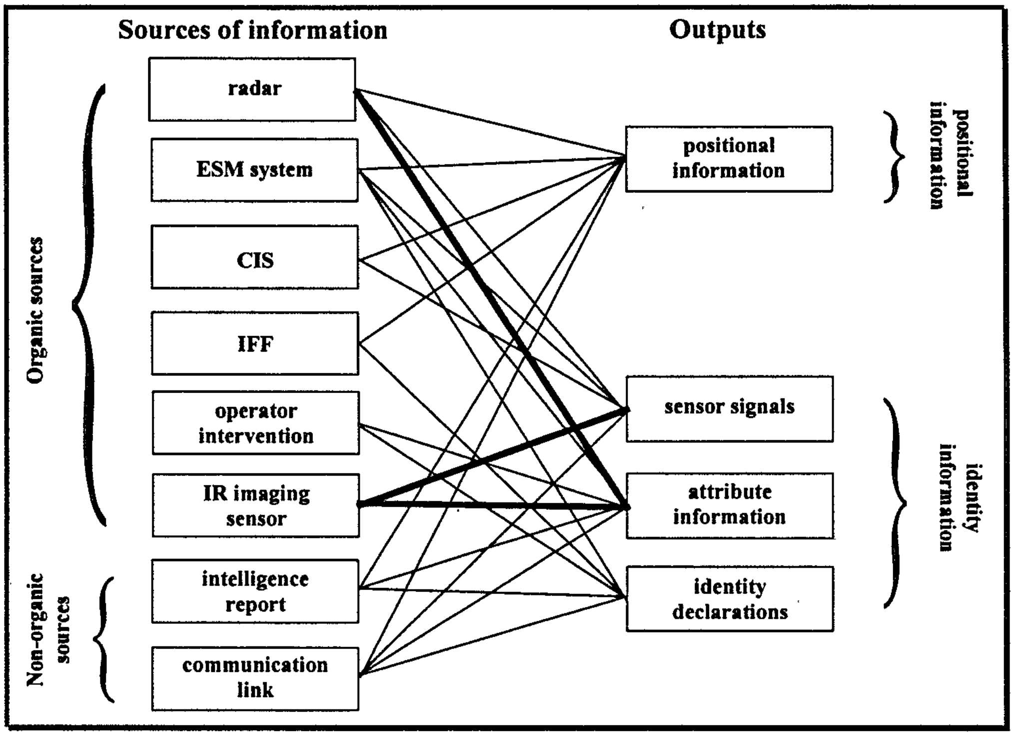


FIGURE 3 - Examples of Advanced Information Sources and Corresponding Outputs

The signal processing unit is responsible for the detection of objects and assessment of the contact's position and attribute information. Again the output of this stage can be sent to a fusion processor. This type of fusion is known as *central level* architecture, where attribute information are combined. The procedure is as follows: a series of target attributes are extracted from sensor measurements, then an inference is drawn between attributes and target types known to possess certain attributes. The feasibility of such an endeavor was shown by Donker (Ref. 15), as well as by Bégin *et al.* (Ref. 16), in the specific case of a Canadian Patrol Frigate. Central level architecture was also demonstrated for an updated frigate in an AWW (Above Water Warfare) environment by Paramax (Ref. 7) and Simard *et al.* (Ref. 17).

Finally, in the target processing unit, the sensor processes many contacts related to the same object and estimates information to produce identity declarations. This information is then sent to the fusion processor to be combined with other identity declarations. This procedure is often called *sensor level* architecture. The following architecture terminology is also used in the literature: data level, feature level and decision level fusion architectures (Ref. 6).

Figure 4, adapted from Paramax (Ref. 7), displays the three fusion architectures providing the structural basis for fusing identity information in the case of 2 sensors. It should be noted that the output from each level of processing represents the three categories of identity information depicted in Fig. 1.

An advantage of the signal level architecture is the minimum information loss incurred since the sensor data are fused directly without approximation via attributes or identity declarations (Ref. 6). However, only data from identical sensors can be fused. Also, its computational requirements are very high due to the complex techniques required to fuse signals. The central level approach results in an information loss from the sensors since sensor data are represented by attributes. Nevertheless, Refs. (7, 16-17) have demonstrated the potential of the central level approach. The sensor level fusion architecture provides a significant loss compared to signal fusion, since data are represented via identity declarations. The information processing for each sensor may thus result in a locally rather than a globally optimized solution, because the fusion process only combines local decisions in the hope of obtaining the correct identity.

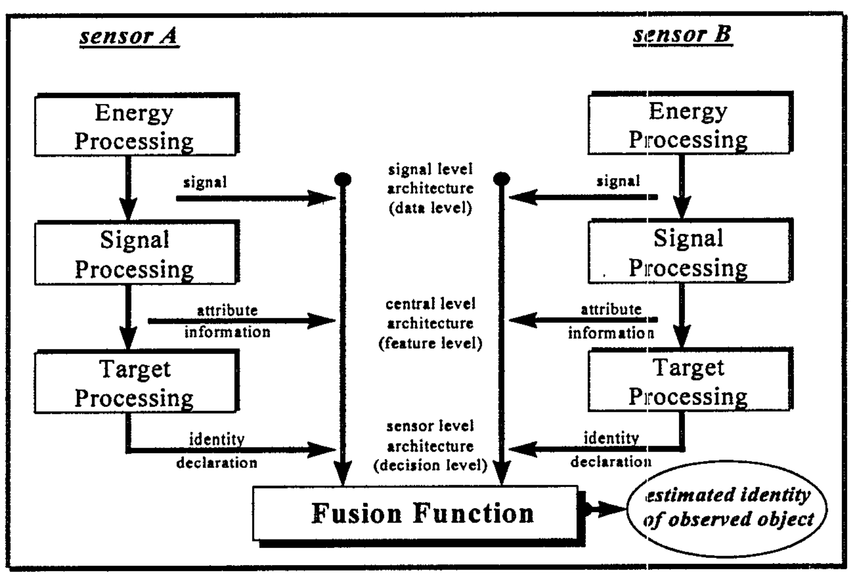


FIGURE 4 - Identity Information Fusion Architectures

Nevertheless, the sensor level architecture approach would seem appropriate if the available identity information is only provided by self-contained sensors producing identity declarations in an autonomous fashion, or by non-organic sources offering identity declarations. If one tentatively assumes that this last scenario is viable, studies are necessary to explore the possibility of combining diverse identity declarations to obtain an estimate of the identity of objects surrounding one or many frigate-type ships. Nahim & Pokoski (Ref. 18), Bogler (Ref. 19), Buede *et al.* (Ref. 20) and Hong & Lynch (Ref. 21) have shown the quantitative benefits of fusing identity information, more specifically identity declarations, in examples with limited scope in which less than 10 different identity declarations were combined. The gains translated into an increased level of performance in identifying objects using the sensor level architecture. In this document, we propose an approach which offers the capability of combining identity declarations pertaining to an AWW environment. The focus of our work will thus be on the **fusion of identity declarations.**

2.3 Fusion of Identity Declarations

To look at the problem of combining identity declarations in a rigorous manner, the following issues need to be addressed: what identity declarations should be combined, how should they be combined and which identity declaration best supports the combined evidence?

2.3.1 What Identity Declarations to Combine?

As depicted in Section 2.1, the quantity of identity declarations available may become quite imposing and diverse. A potential way of organizing identity declarations might be to use the NATO standards for representing maritime tactical information (STANAG 4420, Ref. 22). These standards are in the form of charts that define the full range of tactical information required by the operational user at the command level. The charts were created to establish the basis for developing a standardized representation of spatially displayed tactical data using symbology and colour for NATO maritime units.

Figures 5a and 5b offer an adapted version of a tactical information hierarchy for surface and air objects respectively; certain levels and entries have been omitted for simplicity. The names in the various boxes, which are indicative of the taxonomy used in STANAG 4420, represent identity declaration entities. These entities illustrate the sort of identity declarations provided by various sources which must be combined in order to obtain an estimate of an object's identity. The underlined elements symbolize identity declaration domains; they do not belong to the STANAG 4420 charts but were added to establish a relationship between the domains and their specific entities.

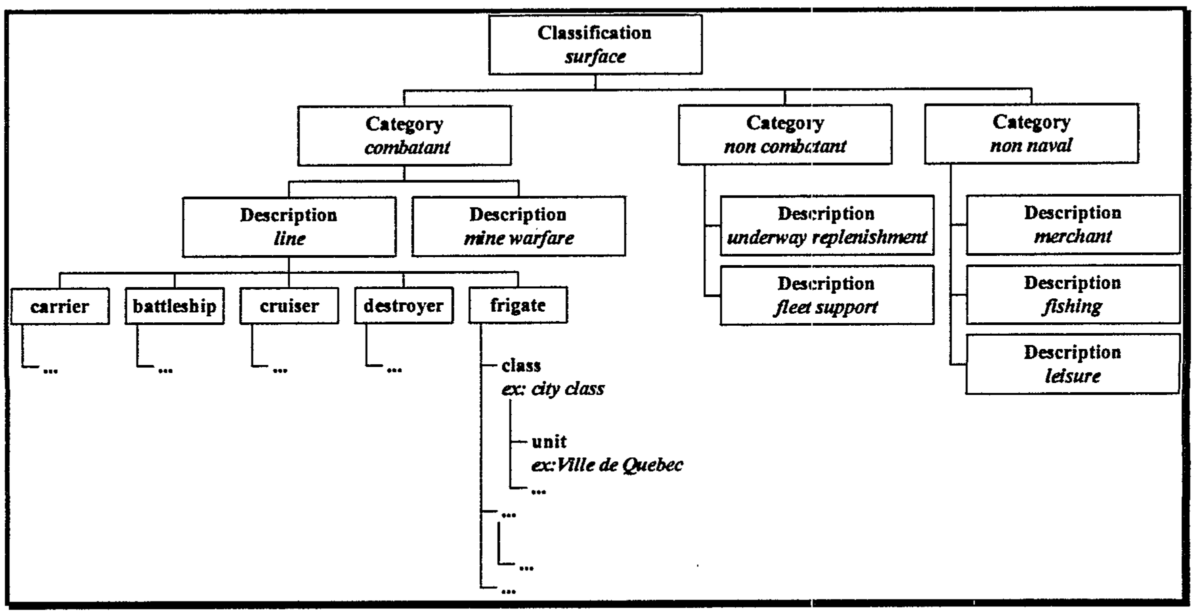


FIGURE 5a - Hierarchy of Tactical Information - Surface Classification

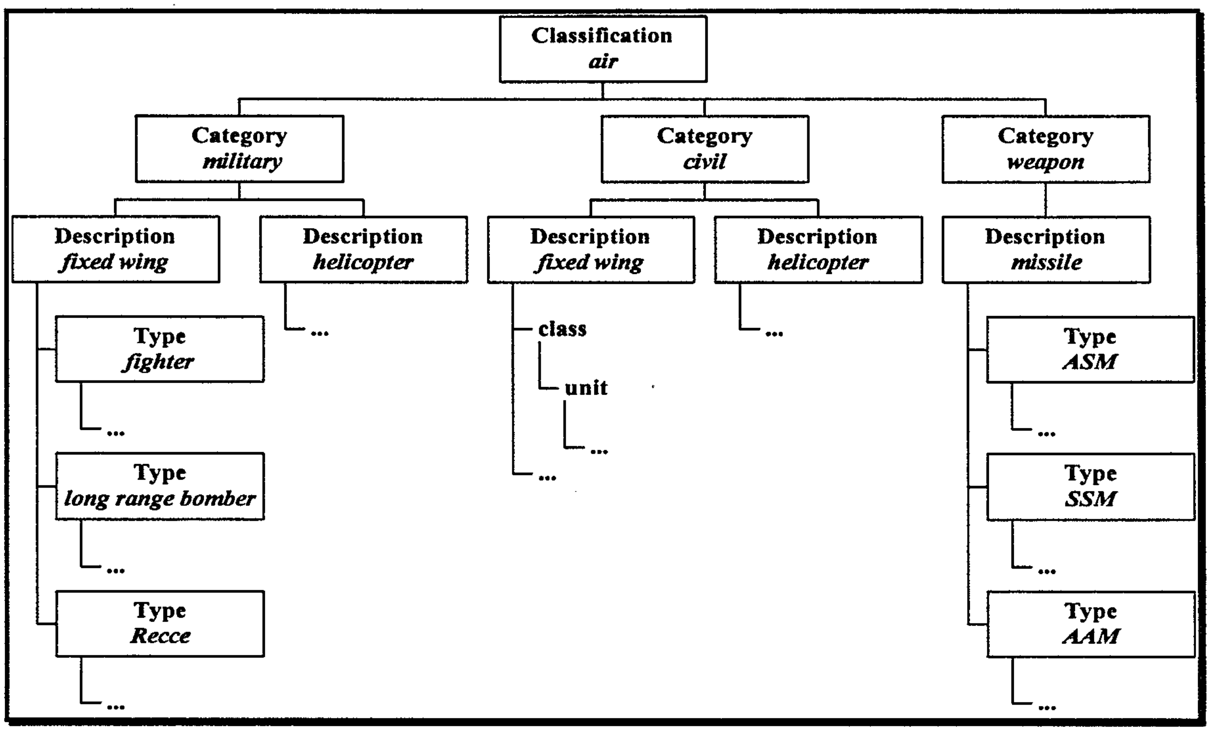


FIGURE 5b - Hierarchy of Tactical Information - Air Classification

Some entities can be further divided into class and unit. These categories typify the identity of objects at two levels of specificity. The "unit" declaration domain characterizes uniquely a detected object.

Also included in the hierarchy of tactical information is the threat designation; its subdivisions are given in Fig. 6.

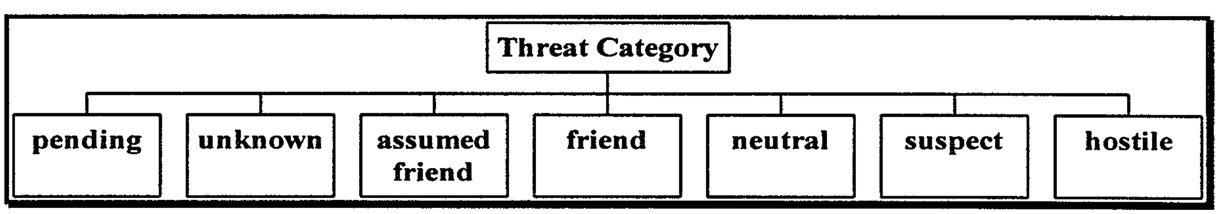


FIGURE 6 - Hierarchy of Tactical Information - Threat Category

The hierarchy of tactical information delineated by Figs. 5a, 5b and 6 should encompass many of the identity declarations provided by the organic and non-organic sources of Fig. 3. Furthermore, because this hierarchy is a NATO standard, it provides a better base for achieving interoperability in information exchange between nations than uncontrolled alternatives.

2.3.2 How to Combine Identity Declarations?

Given that a hierarchical structure has been established between various possible identity declarations, it would be of interest to combine identity declarations pertaining to the same object but provided by various sources in order to obtain a better estimate of the object's identity. Before selecting appropriate combination methods, certain issues need to be discussed such as sensor level fusion architecture and information uncertainty.

2.3.2.1 Sensor Level Fusion Architecture

According to Section 2.2, an appropriate approach to fusing identity declarations is to apply the sensor level architecture. This architecture can be adapted to include identity declarations inferred by operators as well as the ones deduced by non-organic systems.

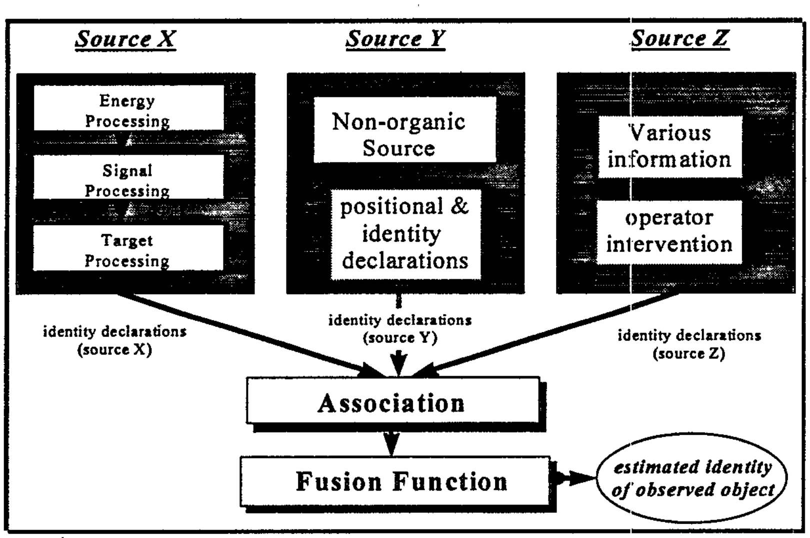


FIGURE 7 - Identity Declaration Fusion Process - Sensor Level Architecture

Figure 7 provides a simple view of the sensor level fusion architecture, adapted from Refs. 6-7. In this approach, each source infers identity declarations independently and these inferences are then combined to obtain an estimated identity of the observed object. Figure 7 shows three dissimilar information sources capable of providing identity declarations. In source X, an organic sensor, the raw input is transformed by the energy processing unit, followed by the signal processing unit to obtain attribute information, and finally by the target processing unit that estimates identity declarations, as delineated in Fig. 4. Non-organic sources such as source Y may directly provide identity declarations compatible with the hierarchy of tactical information. Source Z may produce information concerning the identity of objects, though operator intervention may be necessary to infer declarations.

An important feature of this fusion architecture is the association process. As mentioned earlier, identity of an object is meaningless unless it can be associated with a position. Therefore, identity declarations provided by various independent sources must be partitioned into groups representing observations belonging to the same observed object. Algorithms using positional and identity information must be applied to associate declarations pertaining to the same object but originating from different sources.

As for the main process, called the Fusion Function, it must combine the various identity declarations into a single estimated identity declaration.

2.3.2.2 Identity Declarations: Probabilistic Information

The target processing unit of the sensor level architecture is able to estimate identity declarations by comparing the sensor attribute information with an *a priori* sensor-specific database. This intelligence database may include both identity and kinematic target feature parameters. The matching process of the target processing unit is uncertain due to random type measurement errors and inference errors. Measurement errors may be smoothed out by a signal processing unit as updating is performed. Updating is a process whereby consecutive data from the same sensor, sufficiently separated in time or frequency to be independent, are combined to produce more reliable data. Inference errors occur when the target processing unit infers identity declarations in a form specific to the sensor's domain; an exact match between attribute information and a specific element within the sensor's database is quite unlikely. The inference can often be of low confidence due to the incompleteness of the data used in the process. As an example, Smith & Goggans (Ref. 10) state that there are at least two factors that can render information incomplete in radars. First, the signal-to-noise ratio (SNR) is sometimes insufficient at a given processing gain to obtain the desired measurement. Second, even in the absence of noise, the coded signal does not provide sufficient information to allow a deterministic solution to the problem. Visual observations provide another typical example of uncertain information. There is some level of uncertainty associated with this kind of identification due to the fact that the object may be partially obscured by fog, cloud or darkness, or may simply be too far away to make a conclusive identification (Ref. 8).

For the purpose of this study, we assume that all sources capable of providing identity declarations will do so by attaching to each declaration a quantitative measure of uncertainty, such as 'declaration: frigate, measure of uncertainty: 0.6'. As discussed earlier, the measure of uncertainty in the case of sensors is associated with the measurement and inference errors of the target processing unit. This measure corresponds to the probability that the identity declaration and detected object are the same or, equivalently, to the probability that declaration *i* from source *s* is true:

$$
\begin{aligned}
C_{s,i} &= \text{P(declaration i from source s matches detected object)} \\
&= \text{P(detected object is i, given that source s declared it to be i)} \\
&= \text{P(declaration i from source s is true)}
\end{aligned}
$$

In the case of non-sensor information sources, the matching coefficient $C_{s,i}$ simply typifies a subjective confidence appraisal of the declaration.

2.3.2.3 Uncertainty Fusion Techniques

Whatever options we choose for combining identity declarations, uncertainty techniques are necessary. The problem of combining identity declarations is simply one of fusing uncertain identity declarations rather than one of inferring identity declarations from uncertain information; inference was accomplished by the various sources. Waltz &

Llinas (Ref. 4) propose a taxonomy of identity fusion algorithms whereby some methods infer and others combine identity declarations.

For the purpose of representing and combining uncertain information, many approaches are available. They can be divided into two categories: numerical and non-numerical methods. Among the numerical methods, we find the Bayesian paradigm of probability theory, the certainty factor approach, the Dempster-Shafer theory of evidence, the possibility theory, Thomopoulos's generalized evidence processing theory (GEP), etc. Examples of non-numeric techniques are the theory of endorsements, reasoned assumption approach, non-monotonic logic, etc. A review of numeric and non-numeric methods for handling uncertain information is found in Bhatnagar & Kanal (Ref. 23).

Because we are assuming that identity declarations from various information sources are independent from each other and because an inference method is not required here, the Bayesian paradigm of probability theory (Pearl, Ref. 2) and the Dempster-Shafer evidential theory (Shafer, Ref. 24) are appealing approaches to combining identity declarations. The first is one of the oldest among all numeric approaches to uncertainty. The second, which was conceived as a generalization of the first, is well documented in the literature. Both offer simplified algorithms for combining hierarchically structured information (Refs. 2-3). These two methods will be described, in turn, in Chapters 3 and 4 below.

2.3.3 Decision Criteria

Once identity declarations have been fused, a decision making technique is required to select the identity declaration best supported by the combined evidence (Barnett, Ref. 25). Chapter 5 reviews various approaches that can be used in the face of knowledge combined by the Dempster-Shafer theory. Decision making will then be studied pertaining to information structured in a hierarchical manner.

Chapter 6 describes the identity declaration fusion function based on the findings of Chapters 2 to 5 and provides an example of application in a naval context.

3.0 THE BAYESIAN PARADIGM OF PROBABILITY THEORY

The aim of Bayesian probability theory is to provide a coherent account of how belief should change in light of partial or uncertain information. It is thus an ideal vehicle for representing and combining uncertain information. Judea Pearl (Ref. 2, p. 29) gives a good definition of the Bayesian method:

> The Bayesian method provides a formalism for reasoning about partial beliefs under conditions of uncertainty. In this formalism, propositions are given numerical parameters signifying the degree of belief accorded them under some body of knowledge, and the parameters are combined and manipulated according to the rules of probability.

Before describing the representation and combination rule of the Bayesian method, a review of the axiomatic specifications of probability theory will be given, followed by two possible interpretations of probability.

3.1 Axioms of Probability

The modern axiomatic theory of probability is due to Kolmogorov (Refs. 26-27).

Let us consider a probability space (Ω,Π,P) where:

- Ω is a set, called *sample space*, listing all possible distinct outcomes that may result when a particular experiment is performed.
- Π is a σ - field of subsets of Ω whose elements $A \quad \Pi$ are called events. To be a σ - field, Π must satisfy the following conditions:

  a. Ω is a member of Π,

  b. Π is closed under complementation, i.e.,

  $$A \quad \Pi \Rightarrow \overline{A} \quad \Pi, \text{ where } \overline{A} = \Omega - A$$

  c. Π is closed under countable unions, i.e.

  $$A_1, A_2, \ldots, A_n, \ldots \quad \Pi \Rightarrow \bigcup_{i=1}^{\infty} A_i \quad \Pi.$$

- P is a probability function or probability measure if it assigns a number P(A) to each event A Π in such a way that:
  - a. $P(A) \geq 0$ for all A Π.
  - b. $P(\Omega) = 1$.
  - c. If the A_i are pairwise disjoint members of Π, then

    $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, known as countable additivity.

The above properties of a probability measure P are called Kolmogorov's axioms of probability. The following basic consequences are direct results of these axioms:

$$\forall A \quad \Pi, 0 \leq P(A) \leq 1.$$
$$\forall A \quad \Pi, P(A) + P(\overline{A}) = 1. \tag{3.1}$$

These axioms clearly delineate the constraints of the probability measure P. They are nevertheless capable of supporting various interpretations of probability. The following section gives two particular definitions of probability and shows how each is supported by Kolmogorov's axioms.

3.2 Definitions of Probability

The first definition of probability that comes to mind is the one based on relative frequency. According to that definition, the probability P(A) of an event A stands for the proportion of times that this event occurred in a large (ideally infinite) number of independent trials carried out under identical experimental conditions. A common example is the roll of a balanced dice where the sample space is $\Omega = \{1,2,3,4,5,6\}$ and the field of subsets Π is generated by the elementary events {1}, {2}, {3}, {4}, {5}, {6}. The probability of obtaining number 1 (event {1}) is considered to be 1/6 because the relative frequency of number 1 should approach 1/6 when the dice is rolled a large number of times under similar conditions. In a similar fashion, the probability that numbers 1 or 2 appear should be approximately 2/6. It is important to note that, in practice, the actual observed frequency of, say, the event {1, 2} will vary according to the number of times the dice is rolled and the similarity of the experimental conditions. The frequency interpretation of probability applies only to problems in which there can be a large number of similar repetitions of a certain process (Ref. 28). Because the frequency interpretation of probability satisfies the axioms of probability as delineated by the roll of a dice example, probabilities can be viewed, where appropriate, as statistical measures of proportions. This definition is known as the empirical interpretation of probabilities (Ref. 29); it is also called *objective probability*.

The second approach to probability views it as a degree or measure of belief of an individual about the outcome of an experiment. As mentioned by Bacchus (Ref. 29), probability then becomes an epistemic concept, related to an agent's beliefs, instead of an empirical property related to relative frequency. In this case, probabilities are *subjective* because different individuals may assign different probabilities to the same event, possibly because their information bases are different. However, a single individual should assign the same value to the same event (Ref. 30). For example an individual might want to quantify his/her beliefs concerning the probability of obtaining number 1 from the roll of a dice; the sample space is again $\Omega = \{1,2,3,4,5,6\}$. According to the individual's beliefs, experience and information about the experimental conditions, he/she could assign 1/3 to the likelihood of obtaining number 1. Conversely, the likelihood of not rolling a 1 would be 2/3. Inasmuch as a person's judgments of the relative likelihoods of various combinations of outcomes satisfy the required consistency for his/her subjective probabilities of the different possible events to be uniquely determined, the subjective interpretation of probability satisfies the axioms of probability (Ref. 28).

In our study of identity declaration fusion, both objective and subjective evidence is available. Objective evidence arises from sensor measurements of features which are likely to be corrupted by noise and other distortions (Subsection 2.3.2.2). Therefore the description of the expected feature measurements for known objects can be incomplete or imprecise. Subjective evidence or judgment regards the occurrence or existence of a particular object in the coverage area, based upon *a priori* or expert opinion. Information sources such as operator intervention, intelligence reports and communication links can provide subjective evidence (Section 2.1.1). For this reason, both types of evidence will be used in our study, as they apply to the Bayesian approach.

3.3 The Bayesian Paradigm

The Bayesian paradigm is based on three axioms of probability used to describe the degree of belief of a proposition (hypothesis), given a body of knowledge (evidence). It requires a set of prior probabilities to describe the environment. The evidence pertaining to an event is interpreted in light of the prior probabilities. The results of this analysis is a set of posterior probabilities (Ref. 31). Pearl's definition of the Bayesian method, given at the beginning of this chapter, combines the notions of evidence and belief in the following manner: a hypothesis is attributed a degree of belief, say p, given some evidence. Syntactically, this is written as:

$$P(\text{hypothesis} \mid \text{some evidence}) = p$$

Equivalently,

$$P(H \mid E) = p$$

which specifies the belief in hypothesis H under the assumption that evidence E is known. $P(H \mid E)$ is called Bayes conditionalization and is calculated by the following formula:

$$P(H \mid E) = \frac{P(H \cap E)}{P(E)}.$$

In a similar fashion,

$$P(E \mid H) = \frac{P(E \cap H)}{P(H)},$$

and by substitution, we obtain:

$$P(H \mid E) = \frac{P(E \mid H)\, P(H)}{P(E)} \qquad (3.2)$$

This formula represents the combination rule of the Bayesian method. Equation (3.2) states that the belief in hypothesis H upon obtaining evidence E can be calculated as the likelihood, $P(E \mid H)$, that E will materialize given that H is true, multiplied by the previous belief P(H) in H. P(H | E) is often called the *posterior probability* and P(H) the *prior probability*. The denominator P(E) is a constant which simply ensures that:

$$0 \leq P(H \mid E) \leq 1.$$

3.4 Recursive Formulation of Bayes' Formula

Let us introduce:

E_i = evidence: declaration of object of classification type i; i = 1, ..., I

H_j = hypothesis: presence of object of classification type j; j = 1, ..., J

The events $E_1, \ldots, E_I$ are mutually exclusive, meaning that there is no overlap between them. Let us assume that they are also exhaustive, which implies that they completely describe the possible events. Likewise, let us suppose that events $H_1, \ldots, H_J$ are mutually exclusive and exhaustive.

According to Bayes rule (3. 2), we have:

$$P(H_j \mid E_i) = \frac{P(E_i \mid H_j)\,P(H_j)}{P(E_i)} \tag{3.3}$$

where $P(E_i)$ is given by:

$$P(E_i) = \sum_{s=1}^{J} P(E_i \mid H_s)\,P(H_s).$$

Equation 3.3 gives the probability that the object present is of classification type j given that an object of classification type i was declared. The likelihood matrix $P(E_i \mid H_j)$ can be obtained by allowing the source to observe the objects of interest and make its declaration often enough to obtain a representative sample (Ref. 1). The number of elements in the column must be equal to the number of possible declarations. Therefore,

$$\sum_{i=1}^{I} P(E_i \mid H_j) = 1, \forall j \in \{1, ..., J\}.$$

Schematically, the format of the likelihood matrix is given in Table I in the special case where I = J = 2.

TABLE I
Format of the Likelihood Matrix

		Object present	
		H_1	H_2
Declaration	E_1	$P(E_1 \mid H_1)$	$P(E_1 \mid H_2)$
	E_2	$P(E_2 \mid H_1)$	$P(E_2 \mid H_2)$
		$\Sigma=1.$	

For example, let the likelihood matrix of a given source be:

$$P(E_i|H_j) = \begin{pmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{pmatrix}$$

If an object of classification type 2, H_2, is presented to the source, then E_2 will be the source response 60 percent of the time, while E_1 will be the source response 40 percent of the time. A "perfect" source would be described by the identity matrix (Ref. 18).

Let us now assume that the source can produce, in a time sequential manner, additional information concerning the object in the coverage area. Let E_i^1 and E_k^2 be declarations from the source at time 1 and 2. Therefore the conditional probability that H_j occurs is assessed by taking into account E_i^1 and E_k^2:

$$P(H_j \mid E_k^2 \cap E_i^1) = P(H_j \mid E_k^2, E_i^1) = \frac{P(E_k^2 \cap E_i^1 \mid H_j)\, P(H_j)}{P(E_k^2 \cap E_i^1)} \tag{3.4}$$

If we assume that E_i^1 and E_k^2 are conditionally independent given H_j, then all received information can be combined by simple multiplication (Ref. 32). Equation 3.4 can be rewritten as follows:

$$P(H_j \mid E_k^2, E_i^1) = \frac{P(E_k^2 \mid H_j)\, P(E_i^1 \mid H_j)\, P(H_j)}{P(E_k^2 \mid E_i^1) P(E_i^1)} \tag{3.5}$$

where

$$\begin{aligned} P(E_k^2 \mid E_i^1) &= \sum_{s=1}^{J} P(E_k^2 \cap H_s \mid E_i^1) \\ &= \sum_{s=1}^{J} P(E_k^2 \mid E_i^1 \cap H_s) P(H_s \mid E_i^1) \\ &= \sum_{s=1}^{J} P(E_k^2 \mid H_s) P(H_s \mid E_i^1). \end{aligned}$$

Since $P(E_k^2 \mid H_s) = P(E_k^2 \mid E_i^1 \cap H_s)$ because of conditional independence, substituting $P(H_j|E_i^1)$ for $\dfrac{P(E_i^1 \mid H_j)\, P(H_j)}{P(E_i^1)}$ in (3.5), we obtain:

$$P(H_j \mid E_i^2, E_i^1) = \frac{P(E_k^2 \mid H_j)\, P(H_j \mid E_i^1)}{P(E_k^2|E_i^1)} = \frac{P(E_k^2 \mid H_j)\, P(H_j \mid E_i^1)}{\sum_{s=1}^{J} P(E_k^2 \mid H_s) P(H_s \mid E_i^1)}$$

where $P(H_j \mid E_i^1)$ is the posterior probability calculated after receiving the first evidence.

More generally, it follows by induction that:

$$P(H_j \mid E_{i_{(t)}}^t, E_{i_{(t-1)}}^{t-1}, \ldots, E_{i_{(1)}}^1, E_{i_{(0)}}^0) = \frac{P(E_{i_{(t)}}^t \mid H_j)\, P(H_j \mid E_{i_{(t-1)}}^{t-1}, \ldots, E_{i_{(1)}}^1, E_{i_{(0)}}^0)}{\sum_{s=1}^{J} P(E_{i_{(t)}}^t \mid H_s)\, P(H_s \mid E_{i_{(t-1)}}^{t-1}, \ldots, E_{i_{(1)}}^1, E_{i_{(0)}}^0)} \tag{3.6}$$

where $P(H_j|E_i^0) = P(H_j)$.

To better appreciate the Bayesian approach to uncertainty in terms of representation and combination of information, two simple applications are given in Appendix A (Section A.1).

3.5 The Bayesian Approach to Hierarchical Evidence

As mentioned in Subsection 2.3.1, identity declarations originating from various sources can be represented in a hierarchical fashion. The examples of Appendix A (section A.1) were rather simple in the sense that the evidence was not hierarchically structured. To accommodate hierarchical evidence using the Bayesian formalism, J. Pearl (Refs. 2, 33) devised a method that calculates the impact of an evidence on the belief of every hypothesis in the hierarchy. The definition of a strict hierarchical tree is given below, followed by the description of the technique suggested by J. Pearl. A numerical example is given in Appendix A (Section A.2).

3.5.1 Strict Hierarchical Tree

Let us first define $\Omega = \{H_1, H_2, \ldots H_n\}$ the collection of possible outcomes or hypotheses known to be mutually exclusive and exhaustive. A collection, Ψ, of subsets of Ω is chosen to represent specific events or sets that are of interest. A *strict hierarchical tree* can be created with the events of Ψ if each set has a unique parent set that contains it. For example let:

a = fixed wing/military,
b = helicopter/military,
c = fixed wing/civil,
d = helicopter/civil,
e = missile,

and $\Omega = \{a, b, c, d, e\}$,
$\Psi = \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{a, b\}, \{c, d\}, \{a, b, c, d, e\}\}$.

The strict hierarchy given by the elements of Ψ is represented by Figure 8, which is itself a subset of Figure 5b. Here, some events of Ψ have been renamed to simplify the terminology; for example the event {fixed wing/military, helicopter/military} is called "military" and the event {fixed wing/military, helicopter/military, fixed wing/civil, helicopter/civil, missile} is renamed "air".

Each set is viewed as a node in the tree. The set Ω or "air" becomes the *root of the tree*, the elementary events or single hypotheses of Ω are the *leaves*, and the intermediate nodes represent the unions of their immediate successors.

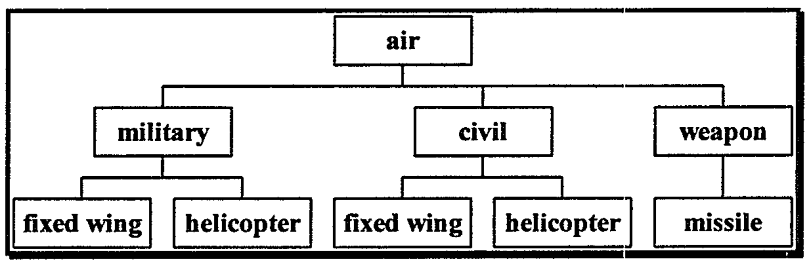


FIGURE 8 - Strict Hierarchical Tree of Hypothesis - Air

3.5.2 Technique Suggested by J. Pearl

The approach first combines the evidence (E), which is in the form of identity declarations, with the current information belonging to the specific hypothesis (H) being updated. Secondly, it distributes this new information to the other hypotheses of the tree according to a proportionality rule. This last step is necessary because the evidence has an impact on the belief of every hypothesis in the hierarchical structure.

The combination of information is performed using a specialization of the Bayesian combination rule to the case of dichotomous alternatives (3.3) (the indices have been omitted to simplify the equations):

$$O(H \mid E) = \lambda_H \, O(H) \tag{3.7}$$

where $\lambda_H = \dfrac{P(E \mid H)}{P(E \mid \overline{H})}$ is the likelihood ratio,

$O(H) = \dfrac{P(H)}{P(\overline{H})}$ is the prior odds, and

$O(H \mid E) = \dfrac{P(H \mid E)}{P(\overline{H} \mid E)}$ is the posterior odds.

Because the evidences are assumed to be conditionally independent, we can obtain from (3.7) a recursive equation similar to (3.6), namely:

$$P(H \mid E^t, E^{t-1}, ..., E^1, E^0) = \alpha_H^t \, \lambda_H \, P(H \mid E^{t-1}, ..., E^1, E^0), \qquad (3.8)$$

where the normalizing factor α_H^t is given by:

$$\alpha_H^t = 1/[\lambda_H \; P(H \mid E^{t-1}, ..., E^1, E^0) + 1 - P(H \mid E^{t-1}, ..., E^1, E^0)] \qquad (3.9)$$

The likelihood ratio λ_H can be regarded as the degree to which the evidence confirms or disconfirms the hypothesis; confirmation is expressed by λ_H>1 and disconfirmation by λ_H<1. The scalar α_H^t effectively normalizes (3.8) since λ_H behaves like a weight; a weight λ_H is given to $P(H \mid E^{t-1}, ..., E^1, E^0)$ and a weight of 1 is given to all other hypotheses belonging to the same hierarchical level.

Once the evidence has been combined with the current hypothesis information using (3.8), the impact at other hierarchical levels must be evaluated. To this end, Pearl (Ref. 2) proposes the following rules, which involve the children of a node H (i.e., the nodes directly below it) and the father of node H (i.e., the node directly above it).

A - Impact on the children of H:

Each child of H will be modified by a factor of $\alpha_H^t \, \lambda_H$. The children's children and so forth will be modified in a similar fashion.

B - Impact on the father F of H:

$$\begin{aligned} P(F \mid E^t) &= \alpha_H^t \, P(F \mid E^{t-1}, ..., E^1, E^0) + P(H \mid E^t, E^{t-1}, ..., E^1, E^0) - \\ &\qquad \alpha_H^t \, P(H \mid E^{t-1}, ..., E^1, E^0) \\ &= \alpha_H^t \, [P(F \mid E^{t-1}, ..., E^1, E^0) - P(H \mid E^{t-1}, ..., E^1, E^0)] + \\ &\qquad P(H \mid E^t, E^{t-1}, ..., E^1, E^0) \end{aligned}$$

The father of the father of H and so forth will be modified in a similar fashion by substituting the appropriate values for hypotheses H and F.

C - Impact on all other hypotheses:

All other hypotheses will be modified by the normalizing factor α^{t}_{H}.

The previous equations obey the rule stating that each node of the tree should acquire a belief equal to the sum of the beliefs belonging to its immediate successors.

Since the combination technique suggested by Pearl is analogous to the combination rule of (3.3), it would have been possible to use (3.8) in the example of Appendix A (Section A.1) if the hypotheses had been hierarchical in nature. Figure 9 provides an example of a strict hierarchical tree of hypotheses: $\Omega = \{C, D, I, K, L, M, N, O, P\}$. As before, other letters are used to represent unions of these outcomes, e.g. $B = \{C, D\}$, $G = \{L, M\}$ and $H = \{K, N, O, P\}$. *A priori* probabilities are indicated for each set of interest. The numerical calculations are presented in Appendix A (Section A.2).

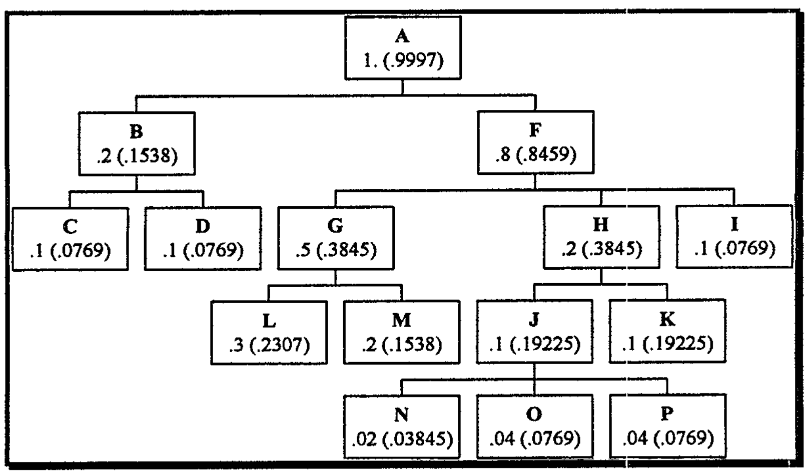


FIGURE 9 - Example of Combination and Propagation of Belief Values Using J. Pearl's Technique

3.6 Comments Concerning the Bayesian Approach

As described earlier, the Bayesian approach requires that all information sources have a complete and accurate knowledge of both the *a priori* probability distribution and the conditional probability matrices. If the sources have little or no knowledge about such things, the Bayesian approach forces them to guess anyway no matter how impoverished the information (Ref. 19). Pearl's updating rules are an example of such guessing activities. In the event where some or all this information is unavailable, this method is at a disadvantage (Ref. 34).

An additional difficulty associated with the Bayesian approach, is the fact that uniform prior probability distributions are often used to represent complete ignorance. For this reason, there is no way of distinguishing between instances of ignorance and instances in which known information suggests a uniform distribution. Thus, if evidence supports proposition (A or B or C) with probability 0.6, it supports individual propositions (A), (B), (C) with probability 0.2. As a result, there is a twofold support of the disjunction of any two of these propositions over the other. In other words, if

$$P(\text{A or B or C}) = 0.6$$

and

$$P(A) = P(B) = P(C) = 0.2$$

then, for example,

$$P(\text{A or B}) = 0.2 + 0.2 = 0.4 = 2 \times P(C).$$

If the probabilities were assigned on the basis of ignorance, however, then there was no evidence to indicate that the disjunct occurrence (A or B) is greater than the singleton of (C). The only proposition supported by the evidence was (A or B or C) and there was no way of distinguishing between subsets of that event (Ref. 35).

Finally, another problem connected with probability coding of beliefs in general is the requirement that the probability of the negation of a hypothesis A, for example, be fixed once the probability of A is known. Because of (3.1), one cannot withhold belief from a proposition without increasing belief in its negation (Ref. 36). However a lack of support for a hypothesis does not necessarily equate to support in its complement. For example, if

$$P(\text{A or B or C}) = .6$$

then, in Bayesian terms:

$$1 - P(\text{A or B or C}) = P(\text{not (A or B or C)}) = .4$$

However, it is important to recognize that if the evidence received is incomplete, then the above implication between (A or B or C) and (not (A or B or C)) cannot be made; the evidence should only support the disjunction, not refute it (Ref. 35).

3.7 Conclusion

Probability theory and the Bayesian paradigm of probability are well formalized methodologies. If *a priori* probability distributions and conditional probability matrices are available and if probabilities can be distributed to single elements, then the Bayesian approach should be used. However, as we have seen above, the Bayesian approach suffers from major deficiencies in a hierarchical context, when fully specified likelihoods are not available. Pearl's rules can be applied but may not always lead to an appropriate estimation of the posterior probabilities associated with certain nodes of a hierarchy. Other problems associated with the Bayesian approach are the coding of ignorance and the strict requirements on the belief of a hypothesis and its negation.

Consequently, the Bayesian approach may not be the ideal technique to fuse uncertain information in the context of identity estimation by multiple dissimilar sources. We will thus investigate a generalization of the Bayesian approach which does not arbitrarily allocate probabilities to the children and parents of a node when this node is updated with new information. This technique is based on the Dempster-Shafer theory of evidence.

4.0 THE DEMSPTER-SHAFER THEORY OF EVIDENCE

The Dempster-Shafer theory was developed by Canadian statistician Arthur Dempster in the 1960's (Ref. 37) and extended by Glenn Shafer in the 1970's (Ref. 24). As in the Bayesian approach, this theory supports the representation of uncertain information and provides a technique for combining it. The idea behind the Dempster-Shafer theory is best described by Shafer himself (Ref. 15):

> The theory of belief functions provides two basic tools to help us make probability judgments: a metaphor that can help us organize probability judgments based on a single item of evidence, and a formal rule for combining probability judgments based on distinct and independent items of evidence.

The Dempster-Shafer technique does not require prior probabilities nor does it need to know the capability of each source. Evidence is not committed to any specific event or set of events until evidence is gained. Also, it is not required that belief not committed to a given proposition should be committed to its negation, nor that belief committed to a given proposition should be committed more specifically (Ref. 15). The technique actually focuses on the probability of a collection of points belonging to the sample space, whereas the classical probability theory is interested in the probability of the individual points (Ref. 38).

The Dempster-Shafer technique has the capability of expressing ignorance explicitly. For example, if A and B are the only hypotheses in the sample space, then $P(A) = P(B) = .5$ indicates that the beliefs in A and B are the same and no ignorance about their occurrences exists. However if the only available information concerns hypothesis A with $P(A) = .5$, it implies that the belief .5 is associated to hypothesis A and the other .5 is attributed to the sample space $\Omega = \{A, B\}$, thereby delineating the ignorance.

The Dempster-Shafer technique does not fully abide by the axioms of probability as stated in Section 3.1. In particular, this technique is not constrained by equation (3.1) but rather supports the following restriction:

$$\forall A \quad \Pi, P(A) + P(\overline{A}) \leq 1. \tag{4.1}$$

Equation 3.1 is obviously a particular case of (4.1), which allows the Bayesian paradigm to be described as a subclass of the theory of evidence.

The following sections describe in a more rigorous fashion the theory behind the Dempster-Shafer evidential approach in terms of representation and combination of evidence.

4.1 Representation of Evidence

4.1.1 Terminology

In the Dempster-Shafer evidential theory, the terminology is slightly different from that used in Probability theory (Ref. 24). The new expressions are in italics. The *frame of discernment* Θ is defined as an exhaustive set of mutually exclusive events or propositions of a particular experiment. It plays the role of the sample space Ω (Section 3.1), so that Θ denotes a set of possible answers to some question where only one answer is correct. We denote 2^{Θ} the set of all subsets of Θ. The elements of 2^{Θ}, or equivalently the subsets of Θ, form the class of general propositions and include the empty set $\varnothing$, which corresponds to a proposition that is known to be false, and the whole set Θ, which corresponds to a proposition that is known to be true. We will assume throughout our discussion that the frame of discernment is finite.

Let A be a subset of Θ. The evidential theory differentiates between the measure of belief committed exactly to A and the total belief committed to A. The first is characterized by the basic probability assignment and the latter by the belief function.

A function m is a *basic probability assignment* if it assigns a number m(A) to each subset A $\quad 2^{\Theta}$, in such a way that:

a. $m(\varnothing) = 0$.

b. $\sum_{A \subseteq \Theta} m(A) = 1.$

The quantity m(A) is called A's *basic probability number* and represents the exact belief in the proposition depicted by A.

A function Bel is a *belief function* if it assigns a number Bel(A) to each subset A 2^{Θ} in such a way that:

a. $Bel(\varnothing) = 0.$
b. $Bel(\Theta) = 1.$
c. for every positive integer n and every collection, $A_1, A_2, \ldots, A_n$ of subsets of Θ,

$$Bel(A_1 \cup \ldots \cup A_n) \geq \sum_i Bel(A_i) - \sum_{i<j} Bel(A_i \cap A_j) + \ldots + (-1)^{n+1} Bel(A_1 \cap \ldots \cap A_n)$$
$$\geq \sum_{\substack{I \subset \{1 \ldots n\} \\ I \neq \varnothing}} (-1)^{|I|+1} Bel(\bigcap_{i \in I} A_i).$$

The quantity Bel(A) is called the *degree of belief* about proposition A. A belief function assigns to each subset of Θ a measure of the total belief in the proposition represented by the subset. Here |I| stands for the cardinality of the set I. The simplest belief function is the one where Bel(Θ) = 1 with Bel(A) = 0 for all A ≠ Θ; this function is called the *vacuous belief function*. A belief function can be obtained from the basic probability assignment:

$$Bel(A) = \sum_{B \subseteq A} m(B), \text{ for all } A \subseteq \Theta. \tag{4.2}$$

A basic probability assignment can in turn be defined as follows with reference to the belief function:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B), \text{ for all } A \subseteq \Theta.$$

In that sense, a basic probability assignment and belief function convey exactly the same information.

A subset A of Θ is called a *focal element* of a belief function Bel over Θ if $m(A) > 0$. The union of all the focal elements of a belief function is called its *core*.

A belief function Bel is called a *simple support function* S if there exists a non-empty subset $A \subseteq \Theta$ and a number $0 \leq s \leq 1$ such that:

$$S(B) = Bel(B) = \begin{cases} 0 \text{ if B does not contain A} \\ s \text{ if B contains A but } B \neq \Theta \\ 1 \text{ if } B = \Theta \end{cases}$$

Therefore, a simple support function precisely supports the subset A and any subset containing A to the degree s, but it provides no support for the subsets of Θ that do not contain A. If S is a simple support function focused on A, then S is the belief function with basic probability numbers:

a. $m(A) = S(A) = s$,
b. $m(\Theta) = 1 - S(A) = 1 - s$,
c. $m(B) = 0$ for all other $B \subseteq \Theta$.

If a simple support function does have a focal element not equal to Θ, then this focal element is called the *focus* of the simple support function.

A *singleton* is a subset of the frame of discernment with only a single member. A simple support function focused on a singleton is a belief function whose only focal elements are the singleton and Θ. If all the focal elements are singletons, then the belief function Bel is called a *Bayesian belief function.*

A belief function is called *dichotomous* with dichotomy $\{A, \overline{A}\}$ if it has no focal elements other than A, $\overline{A}$ and Θ.

Another function, the commonality function, can be obtained from the basic probability assignment. A function Q is a *commonality function* if it assigns a number Q(A) to each subset A 2^{Θ} in such a way that:

a. $Q(\varnothing) = 1$.
b. $\sum_{A \subseteq \Theta} (-1)^{|A|} Q(A) = 0$.

The quantity Q(A) is called A's *commonality number*. It is the sum of basic probability numbers for the set A and all sets which contain it. A commonality function can be defined with reference to the belief function:

$$Q(A) = \sum_{B \subseteq A} (-1)^{|B|} \mathrm{Bel}(\overline{B}), \text{ for all } A \subseteq \Theta \tag{4.3}$$

or with reference to the basic probability assignment:

$$Q(A) = \sum_{A \subseteq B} m(B), \text{ for all } B \subseteq \Theta. \tag{4.4}$$

A belief function can in turn be obtained as follows from the commonality function:

$$\mathrm{Bel}(A) = \sum_{B \subseteq \overline{A}} (-1)^{|B|} Q(B), \text{ for all } A \subseteq \Theta. \tag{4.5}$$

Therefore, the sets of basic probability assignments, belief functions and commonality functions are in one-to-one correspondence and each representation carries the same information as any of the others (Ref. 24).

Yet another function conveys the same information as the belief function; it is called the plausibility function. Let Bel be a belief function over a frame Θ; a function Pl is a *plausibility function* if it assigns a number Pl(A) to each subset A 2^{Θ} in such a way that:

$$\mathrm{Pl}(A) = 1 - \mathrm{Bel}(\overline{A}) \tag{4.6}$$

or

$$\mathrm{Pl}(A) = \sum_{A \cap B \neq \varnothing} m(B), \text{ for all } A \subseteq \Theta \tag{4.7}$$

or

$$\mathrm{Pl}(A) = \sum_{B \subseteq A} (-1)^{|B|+1} Q(B), \text{ for all } A \subseteq \Theta . \tag{4.8}$$

The quantity Pl(A) is known as the *degree of plausibility* of A and expresses the extent to which one fails to doubt A. Pl(A) will be zero when the evidence refutes A, and unity when there is no evidence against A. From equations (4.2) and (4.7) we conclude that:

$$\mathrm{Bel}(A) \leq \mathrm{Pl}(A).$$

The degree of belief and degree of plausibility summarize the impact of the evidence on a particular proposition A in the following manner: the first shows how well the evidence supports proposition A and the second reports on how well its negation $\overline{A}$ is supported. This information can be expressed in the form of an interval called the *evidential interval* whose length, Pl(A) - Bel(A), can be referred to as the *ignorance* remaining about subset A:

$$\text{evidential interval on subset } A = [\mathrm{Bel}(A)\,,\, \mathrm{Pl}(A)].$$

If ignorance about subset A is zero, then the Dempster-Shafer process is identical to the Bayesian approach such that Bel(A) = P(A) = Pl(A); if ignorance about A is equal to one, then no knowledge at all is available concerning subset A.

More generally, the relationship between the Bayesian and Evidential theories can be described by the following equation:

$$\mathrm{Bel}(A) \leq P(A) \leq \mathrm{Pl}(A).$$

Thus, the degree of belief and degree of plausibility on a hypothesis can be seen as the lower bound and upper bound on the probability of that hypothesis. A simple numerical example is presented in Appendix B (Section B.1) to clarify the wealth of terminology associated with the Evidential theory.

4.2 Combination of Evidence

4.2.1 Terminology

As in the Bayesian theory (see equation 3.2), Evidential theory proposes a combination rule, called Dempster's rule of combination, which synthesizes basic probability assignments and yields a new basic probability assignment representing the fused information. The combination rule is known as the *orthogonal sum* and is denoted by $\oplus$. Basically, this rule corresponds to the pooling of evidence: if the belief functions being combined are based on entirely distinct bodies of evidence and the set Θ discerns the relevant interaction between those bodies of evidence, then the orthogonal sum gives degrees of belief that are appropriate on the basis of the combined evidence (Ref. 24).

Let m_1 and m_2 be basic probability assignments, on the same frame of discernment Θ, for belief functions Bel_1 and Bel_2 respectively. If Bel_1's focal elements are $B_1, \ldots, B_k$, and Bel_2's focal elements are $C_1, \ldots, C_n$, the total portion of belief exactly committed to A ($A \neq \varnothing$) is given by the orthogonal sum $m = m_1 \oplus m_2$:

$$m(A) = K \sum_{\substack{i,j \\ B_i \cap C_j = A}} m_1(B_i) \bullet m_2(C_j), \tag{4.9}$$

where

$$1/K = 1 - \sum_{\substack{i,j \\ B_i \cap C_j = \varnothing}} m_1(B_i) \bullet m_2(C_j) = \sum_{\substack{i,j \\ B_i \cap C_j \neq \varnothing}} m_1(B_i) \bullet m_2(C_j). \tag{4.10}$$

The scalar K is a normalizing constant. It normalizes to one the total portion of belief exactly committed to A because it may occur that a focal element B_i of Bel_1 and a focal element C_j of Bel_2 will be such that $B_i \cap C_j = \varnothing$, and

$$\sum_{\substack{i,j \\ B_i \cap C_j = \varnothing}} m_1(B_i) \bullet m_2(C_j) > 0.$$

The recourse to K is thus justified by the need to compensate for the measure of belief committed to $\varnothing$. If $1/K = 0$ then the combined belief functions are said to be totally contradictory and $Bel_1 \oplus Bel_2$ does not exist or, equivalently, Bel_1 and Bel_2 are not

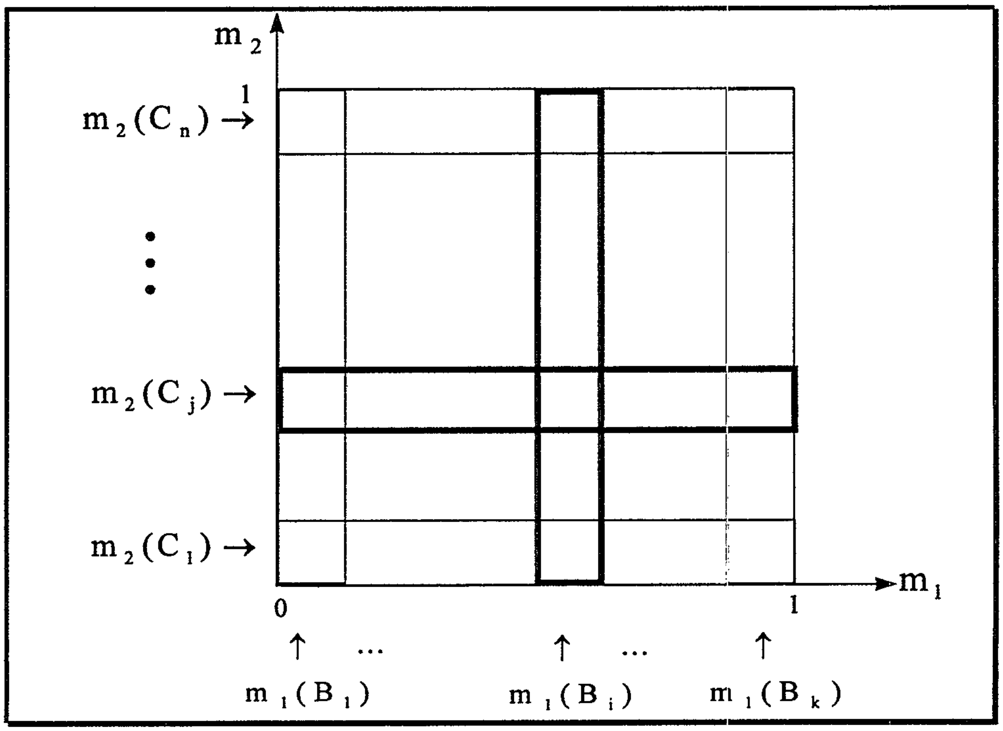


FIGURE 10 - Orthogonal Sum of Two Basic Probability Assignments

combinable. Figure 10 shows the orthogonal sum of two basic probability assignments $m_1(B_i)$ and $m_2(C_j)$; the bold lines delineate the total probability mass committed to $B_i \cap C_j$.

As mentioned earlier, the result of the orthogonal sum is another basic probability assignment; the core of the belief function given by m is equal to the intersection of the cores of Bel_1 and Bel_2.

The operation of orthogonal sum of basic probability assignments satisfies the following properties (Ref. 39):

1. commutativity: $m_1 \oplus m_2 = m_2 \oplus m_1$;
2. associativity: $(m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3)$.

Therefore, the order of combination is immaterial and the operation allows the pairwise composition of a sequence of basic probability assignments such that, if m_1, m_2, ..., m_p are p pieces of evidence, their combination is:

$$m = m_1 \oplus m_2 \oplus \cdots \oplus m_p.$$

If m_i are a collection of basic probability assignments with focal elements A_i, B_j, C_k, D_l, ... respectively, then

$$m(A) = K \sum_{\substack{A_i, B_j, C_k, D_m, \ldots \\ A_i \cap B_j \cap C_k \cap D_m \cap \ldots = A}} [m_1(A_i)\ m_2(B_j)\ m_3(C_k)\ m_4(D_m) \cdots]$$

$$m(\varnothing) = 0 \tag{4.11}$$

$$1/K = 1 - \sum_{A_i \cap B_j \cap C_k \cap D_m \cap \ldots = \varnothing} [m_1(A_i)\ m_2(B_j)\ m_3(C_k)\ m_4(D_m)\ \cdots].$$

It is interesting to note that the formation of orthogonal sums by Dempster's rule corresponds to the multiplication of commonality functions. If the commonality functions for Bel_1, Bel_2 and $Bel_1 \oplus Bel_2$ are denoted by Q_1, Q_2 and Q, respectively, then

$$Q(A) = K\ Q_1(A)\ Q_2(A), \tag{4.12}$$

where K does not depend on A. The proof of (4.12) is given in Appendix B (Section B.2).

It is therefore possible to calculate $Bel_1 \oplus Bel_2$ by applying the following procedure:

a. Find the plausibility functions Pl_1 and Pl_2 from (4.6):

$$Pl_i = 1 - Bel_i(\overline{A}).$$

b. Find the commonality functions Q_1 and Q_2 (from a transformation of (4.8)):

$$Q_i(A) = \sum_{\varnothing \neq B \subseteq A} (-1)^{|B|+1} Pl_i(B). \tag{4.13}$$

c. Find the appropriate normalization constant K by:

c1. Substituting A for Θ in (4.8) such that $Pl(\Theta)=1$,

c2. Substituting Q(B) for $K\ Q_1(B)\ Q_2(B)$ in (4.8),

c3. Evaluating $1/K = \sum_{\varnothing \neq B \subseteq \Theta} (-1)^{|B|+1} Q_1(B)\ Q_2(B)$. (4.14)

d. Find the multiplication of commonality functions using plausibility (from (4.8) and (4.12)):

$$Pl(A) = K \sum_{\varnothing \neq B \subseteq A} (-1)^{|B|+1} Q_1(B)\ Q_2(B). \tag{4.15}$$

e. Find the belief from the resulting plausibility function.

In a similar fashion, the formulas all generalize to the case where more than two belief functions are combined by replacing $Q_1(B)\ Q_2(B)$ by $Q_1(B) \cdots Q_n(B)$. Appendix B (Section B.3) shows a numerical example of Dempster's rule of combination

4.3 Comments Concerning Dempster' s Rule of Combination

Various authors have commented on Dempster's rule of combination in terms of its justification and normalization inadequacies. Some authors have even proposed replacements to his combination rule. These concerns are summarized in the following subsections.

4.3.1 Requirements of Dempster's Rule of Combination

Dempster's rule of combination is simply a rule for computing a belief function from two other belief functions. According to Shafer (Ref. 24), this rule reflects the pooling of evidence within the Dempster-Shafer theory provided that two requirements are met: the bodies of evidence to be combined have to be independent and the frame of discernment must discern the relevant interaction of these bodies of evidence. We will argue that within the framework of the current study, Shafer's requirements are met.

Shafer (Ref. 24) states that Dempster's combination rule seems to reflect the pooling of evidence provided that the belief functions to be combined are actually based on entirely distinct bodies of evidence. He later provides further insight into his independence requirement by suggesting that the evidences to be combined must be independent when viewed abstractly, i.e., before the interactions of their conclusions are taken into account. However, Shafer has not provided a formal definition of independence, such as $A \perp B \Leftrightarrow P(A \cap B) = P(A)\,P(B)$.

Voorbraak (Ref. 40) has shown, using a simple example (given in Appendix B, Section B.4), that even if the evidences seem independent according to Shafer's definition, the combination can produce counterintuitive results.

In the study of identity declaration fusion, if we assume that the information sources are independent from one another according to Shafer's definition, then the independence requirement becomes much simpler, since no inference process is applied to the evidence:

$$\text{id fusion example}\begin{cases}\text{X: declaration from source 1} \Rightarrow \text{A: classification type 1}\\ \text{Y: declaration from source 2} \Rightarrow \text{A: classification type 1}\end{cases}$$

Here also $\mathrm{Bel}_X(A) \oplus \mathrm{Bel}_Y(A)$ will be acceptable, provided that the evidences seem independent according to Shafer's definition. In what follows, we will assume that for the problem of identity declaration fusion, Shafer's independence requirement is met.

The second topic to be investigated is the discernment of evidence. Dempster's rule should only be used if the frame of discernment Θ is fine enough to discern all relevant interaction of the evidence to be combined. Two definitions are in order to better understand the concept of discernment of evidence.

Given finite sets Θ and Ω, a mapping $\omega{:}2^{\Theta} \rightarrow 2^{\Omega}$ is called a *refining* if the sets $\omega(\{\theta\})$ constitute a disjoint partition of Ω:

$$\bigcup_{\theta\in\Theta}\omega(\{\theta\}) = \Omega,$$

and the sets $\omega(\mathrm{A})$ are given in terms of the $\omega(\{\theta\})$:

$$\omega(\mathrm{A}) = \bigcup_{\theta\in\mathrm{A}}\omega(\{\theta\}), \quad \text{for each } \mathrm{A} \subset \Theta.$$

Whenever $\omega{:}2^{\Theta} \rightarrow 2^{\Omega}$ is a refining, Ω is a refinement of Θ, and Θ is a *coarsening* of Ω.

Shafer (Ref. 24) states that Dempster's rule of combination may give inaccurate results when applied in a frame of discernment that is too coarse. Indeed, if S_1 and S_2 are support functions over a frame Θ that are based on distinct bodies of evidence, and if Ω is a *coarsening* of Θ, then Dempster's rule applied in the frame Θ yields the support function:

$$((S_1|2^{\Theta}) \oplus (S_2|2^{\Theta}))|2^{\Omega}.$$

However, if it is applied in the frame Ω, the support function becomes:

$$((S_1|2^{\Omega}) \oplus (S_2|2^{\Omega}))|2^{\Omega}.$$

These two support functions may, in fact, differ. This can clearly be seen in the example of Section B.5 (Appendix B).

4.3.2 Normalization Inadequacies

A controversial issue in the Dempster-Shafer theory relates to the normalization of probabilities and its role in Dempster's rule of combination of evidence (Refs. 41-44). As mentioned in Section 4.2, normalization compensates for the measure of belief committed to the empty set. However, normalization can lead to highly counterintuitive results because it suppresses an important aspect of the evidence. The following theoretical example, adapted from Zadeh (Ref. 42), illustrates this point.

Let Θ be the frame of discernment of possible diagnoses for patient P: {meningitis, brain tumor, concussion}. Suppose that the first observation is: doctor X diagnoses that patient P has either meningitis, with probability 0.99, or brain tumor, with probability 0.01. The second observation is: doctor Y diagnoses that patient P has either concussion, with probability 0.99, or brain tumor, with probability 0.01. Applying Dempster's rule, as shown below, leads to the conclusion that the belief that patient P has brain tumor is 1.0 which is a very unlikely result.

m_2 \ m_1	m_1 ({meningitis}) .99	m_1 ({brain tumor}) .01	m_1 ({concussion}) .0	m_1 (Θ) .0
m_2 ({meningitis}) .0	{meningitis} .0	{} .0	{} .0	{meningitis} .0
m_2 ({brain tumor}) .01	{} .0099	{brain tumor} .0001	{} .0	{brain tumor} .0
m_2 ({concussion}) .99	{} .9801	{} .0099	{concussion} .0	{concussion} .0
m_2 (Θ) .0	{meningitis} .0	{brain tumor} .0	{concussion} .0	Θ .0

<u>Before Normalization</u>

m({meningitis}) = 0.0
m({brain tumor}) = 0.0001
m({concussion}) = 0.0
m(Θ) = 0.0
m(∅) = 0.9999

<u>After Normalization</u> (1/K = 0.0001)

m({meningitis}) = 0.0
m({brain tumor}) = 1.0
m({concussion}) = 0.0
m(Θ) = 0.0
m(∅) = 0.0

Yager (Ref. 39) proposed an alternative rule of combination which does not produce counterintuitive results when there is conflict between pieces of evidence. His combination rule is denoted by $\perp$. Let m_1 and m_2 be basic probability assignments, on the same frame of discernment Θ, for belief functions Bel_1 and Bel_2 respectively. If Bel_1's focal elements are $B_1, \ldots, B_k$, and Bel_2's focal elements are $C_1, \ldots, C_n$, the total portion of belief exactly committed to A $(A \neq \varnothing, \Theta)$ is given by the sum $m = m_1 \perp m_2$:

$$m(A) = \sum_{\substack{i,j \\ B_i \cap C_j = A}} m_1(B_i) \cdot m_2(C_j)$$

$$m(\Theta) = \left(\sum_{\substack{i,j \\ B_i \cap C_j = \Theta}} m_1(B_i) \cdot m_2(C_j) \right) + K$$

$$m(\varnothing) = 0$$

$$K = \sum_{\substack{i,j \\ B_i \cap C_j = \varnothing}} m_1(B_i) \cdot m_2(C_j)$$

The fundamental distinction between this modified combination rule and Dempster's original proposal is that under the latter, via the normalization factor, the belief K, which is the total conflict, is proportionally allocated to the focal elements of m. Therefore the contradictory portion is disregarded. With the use of the new rule, the conflicted portion of the belief is put back into the set Θ, as it regards the contradiction as coming from ignorance. Applying Yager's rule to the diagnosis example, we obtain the following results which seem intuitively more plausible:

<u>Before applying Yager's rule</u>	<u>After applying Yager's rule</u>
$m(\{\text{meningitis}\}) = 0.0$	$m(\{\text{meningitis}\}) = 0.0$
$m(\{\text{brain tumor}\}) = 0.0001$	$m(\{\text{brain tumor}\}) = 0.0001$
$m(\{\text{concussion}\}) = 0.0$	$m(\{\text{concussion}\}) = 0.0$
$m(\Theta) = 0.0$	$m(\Theta) = 0.0 + 0.9999 = 0.9999$
$m(\varnothing) = 0.9999$	$m(\varnothing) = 0.0$

Yager's rule, applied to the example of Appendix B (Section B.3), produces similar results to those obtained by Dempster's rule. However, Inagaki (Ref. 45) notes that Yager's rule is not associative, in that results are dependent on the order in which data are received. This makes Yager's rule considerably less attractive.

Another option proposed by Yager (Ref. 39) is not to modify Dempster's rule, but rather to suggest when the latter should be used in order to avoid undesired results. Its methodology considers the combined basic probability assignment m based on p pieces of evidence to be a good and informative combination if it satisfies the following, rather subjective, conditions:

a. In formulating m, consider all the information available $(m_1, m_2, \ldots, m_p)$.
b. The information used to obtain m must not be highly conflicting.
c. The specificity of m is high.
d. The entropy of m is high.

Specificity and entropy measure the amount of information contained in a basic probability assignment. *Specificity*, P_m, measures the degree to which the basic probability numbers are allocated to focal elements small in size; it provides an indication of the dispersion of the belief. The higher P_m, the less diverse is the evidence. If we assume that m is a belief structure defined over the set X and Θ has cardinality n, then:

$$P_m = \sum_{A \subset X, A \neq \varnothing} m(A) / n_A, \qquad n_A = \text{Card } A. \tag{4.16}$$

This quantity is characterized by the following properties (see proof in Appendix B, Section B.6):

a. $1/n \leq P_m \leq 1$,
b. $P_m = 1/n$ iff m is vacuous,
c. $P_m = 1$ iff m is a Bayesian belief function.

Entropy, E_m, measures the degree to which the basic probability numbers are allocated in a consonant manner, that is, not allocating mass among disjoint sets. It thus provides a measure of the dissonance of the evidence. The lower E_m, the more consistent is the evidence. If m is a belief structure defined over the set X, then the entropy measure is given by:

$$E_m = -\sum_{A \subset X} m(A)\ln(Pl(A)) = \sum_{A \subset X} m(A)\,Con(Bel, Bel_A), \tag{4.17}$$

where $Con(Bel, Bel_A) = -\ln(1-k)$ and $k = \sum_{\substack{i,j \\ A_i \cap B_j = \varnothing}} m_1(A_i)\,m_2(B_j)$.

The entropy measure is characterized by the following properties (proofs in Appendix B, Section B.7):

a. If m is a Bayesian belief function, E_m reduces to the Shannon entropy measure,
b. $0 \le E_m \le \ln(n)$,
c. $E_m = 0$ if $A_i \cap A_j = \varnothing$ for each pair of focal elements,
d. $E_m = \ln(n)$ iff $m(A_i) = \frac{1}{n}$ for i=1, 2,...,n.

Entropy can also be described by H_m, which is a transformation of E_m :

$$H_m = e^{-E_m} = e^{\sum_{A \subset \Theta} \ln[Pl(A)]^{m(A)}} = e^{\ln[Pl(A_1)]^{m(A_1)} + \ln[Pl(A_2)]^{m(A_2)} + \ldots} \text{ for all focal elements } A_i$$

$$= Pl(A_1)^{m(A_1)} \times Pl(A_2)^{m(A_2)} \times \ldots = \prod_{A \subset \Theta} Pl(A)^{m(A)}. \tag{4.18}$$

It can be shown that

a. $1/n \le H_m \le 1$,
b. $H_m = 1$ if $A_i \cap A_j \ne \varnothing$ for each pair of focal elements,
c. $H_m = 1/n$ iff $m\{B_i\} = 1/n$ for i=1, 2,...,n.

Thus the two measures P_m and H_m are such that the closer they are to unity, the more informative the evidence. The *degree of informativeness* of a basic probability assignment m can be obtained from:

$$I_m = P_m H_m \tag{4.19}$$

where $1/n^2 \le I_m \le 1$.

4.4 Computational Complexity

One drawback of the Dempster-Shafer Evidential Theory is the long calculation time required by its high computational complexity. Because the combination rule produces basic probability numbers on the subsets of Θ, the calculations are time exponential. In comparison, the Bayesian approach provides probability statements on the elements of Θ. Therefore, if Θ consists of 4 possible points/outcomes, the definition of a probability function on Θ requires the assignment of probability to 4 points, whereas the definition of the Dempster-Shafer basic probability assignment requires the definition of m(A) for 2^4=16 subsets A of Θ.

Three categories of options are available to reduce computational complexity. The first approximates the belief function, the second treats simple support functions instead of belief functions, and the last one separates the frame of discernment into smaller, more manageable frames, one for each set of mutually exclusive hypotheses. Table II summarizes the various options available for reducing the computational complexity of Dempster's combination rule.

Voorbraak (Ref. 46) has defined a Bayesian approximation of a belief function and he has shown that combining the Bayesian approximations of belief functions is computationally less complex than combining the belief functions themselves; the computational time will be reduced from exponential to polynomial. This approach, however, is only appealing when one is interested in final conclusions about the elements of Θ; in the study of identity declaration fusion, we are mostly interested in subsets of Θ.

J. Barnett (Ref. 47) demonstrated that if each piece of evidence consists of simple support functions focused on singleton propositions and their negations, computational time will be reduced from exponential to linear. Another option was proposed by Gordon & Shortliffe (Ref. 48). It is based on the assumptions that (i) each piece of evidence consists of simple support functions focused for or against subsets of Θ instead of singletons, and that (ii) the subsets of Θ can be structured in a strict hierarchical tree. This method builds on Barnett's approach while permitting hierarchical relationships among hypotheses; its aim is similar to that of Pearl (Section 3.6.2).

TABLE II

Options to Reduce Computational Complexity of Dempster's Rule

Author(s)	Technique	Calculation
Voorbraak (Ref. 46)	Bayesian approximation of belief function	Polynomial time proportional to $\vert\Theta\vert$
Barnett (Ref. 47)	Simple support function focused on singleton	Linear time proportional to $\vert\Theta\vert$
Gordon & Shortliffe (Ref. 48)	Simple support function using subsets of Θ, evidence hierarchically structured	Linear time proportional to $\vert\Theta\vert$
Shafer (Ref. 49)	Belief functions carried by the field of subsets generated by children of node, evidence hierarchically structured	Proportional to size of sibs
Shafer & Logan (Ref. 3)	Simple support function focused on a subset of Θ or its complement, and carried by the field of subsets generated by children of node, evidence hierarchically structured	Linear time proportional to number of nodes in the tree

However, combining negative evidence leads to computational difficulties because the intersection of the complements of nodes may fail to correspond to a node or its complement. In such a case, an approximation is suggested by the authors but this approximation restricts the usefulness of the plausibility measure. The algorithm can be implemented in a form which is linear in the number of nodes in the tree.

Cleckner (Ref. 35) offers a comparative study in terms of memory requirements and computational complexity in which the standard Dempster-Shafer combination rule is compared with Barnett's algorithm and the alternative due to Gordon & Shortliffe (Ref. 47). Cleckner concluded that when dealing with simple support functions focused on singleton hypothesis, Barnett's technique is the least computer intensive of the three.

The last category of options, which also assumes that evidence is hierarchically structured, was proposed by Shafer (Ref. 49). He suggested that the belief functions to be combined should be carried by a partition P of Θ, which has fewer elements than Θ. This is done by separating the frame of discernment into smaller more manageable frames, one for each set of mutually exclusive hypotheses. Of course, once the elements have been separated into multiple frames, items from different frames can no longer be compared since they then pertain to different belief functions. Shafer's first approach combines belief functions each of which is carried by the field of subsets generated by the children of a particular node. His second approach combines simple support functions focused on a subset of Θ or its complement (Shafer & Logan, Ref. 3). The latter uses the same type of evidence as considered by Gordon & Shortliffe (Ref. 48), while avoiding some of its shortcomings. This approach is detailed in the following section.

4.5 Dempster's Rule for Hierarchical Evidence, Revisited by Shafer and Logan

The Shafer & Logan technique reduces computational complexity in three ways:

a. By using hierarchical evidence that reduces the number of admissible subsets;
b. By using simple support functions focused on a subset of Θ or its complement;
c. By reducing Dempster's rule of combination to a series of combinations involving smaller frames of discernment.

The third item is of prime importance. By reducing the frame of discernment Θ into smaller frames of discernment, the complexity is reduced because the smaller frames have less elements than Θ. However, a constraint has to be imposed for the combination to be permissible in the smaller frames: the simple support functions and their combination to be permissible in the smaller frames: the simple support functions and their combination must be carried by the field of subsets generated by the children of a node. To understand this constraint, certain terms will be defined in the following subsections.

4.5.1 Partition of a Frame of Discernment

A *partition* of a frame of discernment Θ is a set of disjoint non-empty subsets of Θ whose union equals Θ; such a partition P can itself be regarded as a frame of discernment. P^* represents the set consisting of all unions of elements of P. For example, if $\Theta = \{a, b, c, d\}$ and $P = \{\pi_1, \pi_2\}$ with $\pi_1 = \{a\}$ and $\pi_2 = \{b, c, d\}$, then $P^* = \{\varnothing, \{a\}, \{b, c, d\}, \{a, b, c, d\}\}$.

As defined in Subsection 4.3.1, partition P_1 is a refinement of partition P_2 if for every element P_1 in P_1 there is an element P_2 in P_2 such that $\mathrm{P}_1 \subseteq \mathrm{P}_2$. For example, let P_1 and P_2 be two partitions of Θ; let also $\pi_1 = \{a\}$, $\pi_2 = \{b, c, d\}$ and $\pi_3 = \{a, b, c, d\}$. If $P_1 = \{\pi_1, \pi_2\}$, $P_1^* = \{\varnothing, \{a\}, \{b, c, d\}, \{a, b, c, d\}\}$ and $P_2 = \{\pi_3\}$, $P_2^* = \{\varnothing, \{a, b, c, d\}\}$, then P_1 is a refinement of P_2.

If all the subsets of a belief function Bel are included in P^*, the belief function Bel is then said to be *carried by P*. For example, let $P = \{\pi_1, \pi_2\}$ with $\pi_1 = \{a\}$ and $\pi_2 = \{b, c, d\}$ and the belief function Bel be represented by the basic probability assignments:

$$m(\{a, b\}) = 0.5$$

$$m(\{a, b, c, d\}) = 0.5.$$

The belief function Bel is not carried by P since the subset $\{a,b\}$ P^* where the latter is defined as before.

Shafer & Logan have shown that if Bel_1 and Bel_2 are both carried by P, then $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$ will also be carried by P. This result has an important impact on the combination of evidence when using, for example, commonality functions as described in Subsection 4.2.1. In effect, this conclusion may be transposed to (4.13), (4.14) and (4.15) if it is assumed that Bel_1 and Bel_2 are both carried by P and that $A \in P^*$. One finds:

$$Q_i(A) = \sum_{\substack{B \in P^* \\ \varnothing \neq B \subseteq A}} (-1)^{|B|^P+1} Pl_i(B)$$

$$1/K = \sum_{\varnothing \neq B \in P^*} (-1)^{|B|^P+1} Q_1(B)\, Q_2(B) \tag{4.20}$$

$$Pl(A) = K \sum_{\substack{B \in P^* \\ \varnothing \neq B \subseteq A}} (-1)^{|B|^P+1} Q_1(B)\, Q_2(B)$$

where $|B|^P$ denotes the number of elements of P contained in B. The end result is that the plausibility function Pl for $Bel_1 \oplus Bel_2$ can be computed, because the combination is carried by P.

To say that $Bel_1 \oplus Bel_2$ is carried by P is equivalent to saying that P discerns the interaction between Bel_1 and Bel_2. The topic of discernment of evidence was introduced in Subsection 4.3.1. We say that P discerns the interaction relevant to itself if:

$$(Bel_1|2^P) \oplus (Bel_2|2^P) = (Bel_1 \oplus Bel_2)|2^P$$

4.5.2 General Concepts

Shafer & Logan (Ref. 3) have introduced a new terminology and notation that will greatly facilitate the understanding of the combination process.

Let $\mathfrak{R}$ be the collection of all nodes below Θ. As illustrated in Figure 11, $\mathfrak{R}$ is represented by {C, D, E, F, G, H, I, J, K}. As in Subsection 3.6.2, D is said to be the child of C, and C is D's parent. The set of nodes that consists of all the children of a given non terminal node is called a *sib*; the sib ℓ_C consists of the children of C.

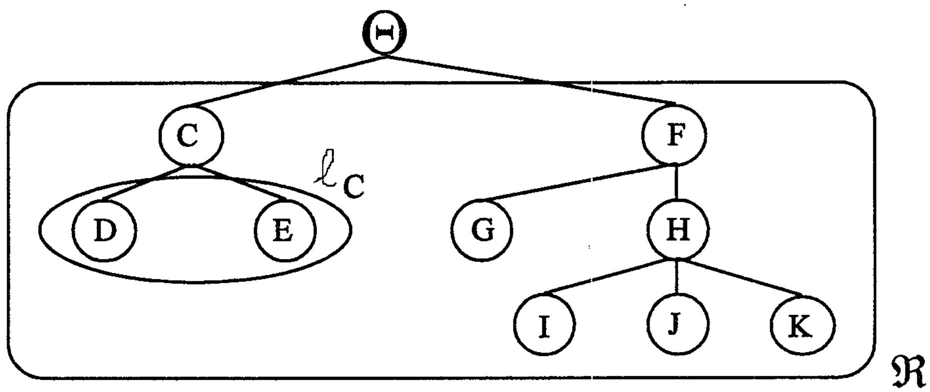


FIGURE 11 - Hierarchical Evidence to Illustrate $\Re$ and ℓ_C

It is assumed that for each node A in $\Re$, there is a single dichotomous belief function Bel_A with dichotomy $\{A, \overline{A}\}$. Furthermore, $\mathrm{Bel}_A(A)$ and $\mathrm{Bel}_A(\overline{A})$ are both strictly less than one, but either, or both, can be zero.

For each node A in the tree, $\mathrm{Bel}_A^{\downarrow}$ denotes the orthogonal sum of Bel_B for all nodes B that are strictly below A.

For each node in $\Re$, $\mathrm{Bel}_A^{\diamond}$ denotes the orthogonal sum of Bel_B for all nodes B in $\Re$ that are neither below A nor equal to A.

For example, from Figure 11:

$$\begin{aligned}\mathrm{Bel}_F^{\downarrow} &= \mathrm{Bel}_G \oplus \mathrm{Bel}_H \oplus \mathrm{Bel}_H^{\downarrow} \\ &= \mathrm{Bel}_G \oplus \mathrm{Bel}_H \oplus \mathrm{Bel}_I \oplus \mathrm{Bel}_J \oplus \mathrm{Bel}_K\end{aligned}$$

and

$$\mathrm{Bel}_C^{\diamond} = \mathrm{Bel}_F \oplus \mathrm{Bel}_F^{\downarrow}.$$

If we generalize to Θ, we have

$$\mathrm{Bel}_\Theta^{\downarrow} = \mathrm{Bel}_C^{\downarrow} \oplus \mathrm{Bel}_C \oplus \mathrm{Bel}_C^{\diamond}$$

or, equivalently

$$\mathrm{Bel}_\Theta^{\downarrow} = \mathrm{Bel}_F^{\downarrow} \oplus \mathrm{Bel}_F \oplus \mathrm{Bel}_F^{\diamond}.$$

In terms of this nomenclature, the aim of Shafer & Logan's approach amounts to calculating the values of all the nodes $A \in \Re$:

$$\mathrm{Bel}_{\Theta}^{\downarrow} = \oplus\{\mathrm{Bel}_{A} | A \in \Re\},$$

or, equivalently

$$\mathrm{Bel}_{\Theta}^{\downarrow} = \mathrm{Bel}_{A}^{\downarrow} \oplus \mathrm{Bel}_{A} \oplus \mathrm{Bel}_{A}^{\Diamond}. \tag{4.21}$$

Using the definitions of partition of a frame of discernment and the concept of discernment of evidence from Subsection 4.5.1, Shafer & Logan have shown that:

a. if P is a partition of Θ and $P \in \Re \cap P$, then $(\mathrm{Bel}_{P}^{\downarrow})_{P} = (\mathrm{Bel}_{P}^{\downarrow})_{\{P,\bar{P}\}}$,

b. if P is a partition of Θ, $A \in \Re$ and $\overline{A} \in P$ then $(\mathrm{Bel}_{A}^{\Diamond})_{P} = (\mathrm{Bel}_{A}^{\Diamond})_{\{A,\bar{A}\}}$,

c. if P is a partition of Θ, then P discerns the interaction relevant to itself among the belief functions $\{\mathrm{Bel}_{P}^{\downarrow} | P \in \Re \cap P\}$ and $\{\mathrm{Bel}_{P} | P \in \Re \cap P\}$.

From the characteristics of a partition of a frame of discernment and the definition of $\mathrm{Bel}_{A}^{\downarrow}$, we obtain, say, for a partition $\ell_{A} \cup \{\overline{A}\}$:

$$(\mathrm{Bel}_{A}^{\downarrow})_{\ell_{A} \cup \{\bar{A}\}} = \oplus\{(\mathrm{Bel}_{B})_{\ell_{A} \cup \{\bar{A}\}} \oplus (\mathrm{Bel}_{B}^{\downarrow})_{\ell_{A} \cup \{\bar{A}\}} | B \in \ell_{A}\}.$$

By a. above, this is equivalent to

$$(\mathrm{Bel}_{A}^{\downarrow})_{\ell_{A} \cup \{\bar{A}\}} = \oplus\{\mathrm{Bel}_{B} \oplus (\mathrm{Bel}_{B}^{\downarrow})_{\{B,\bar{B}\}} | B \in \ell_{A}\}. \tag{4.22}$$

Note that if the element B of ℓ_{A} is a terminal node, then $\mathrm{Bel}_{B}^{\downarrow}$ is vacuous and the orthogonal sum $\mathrm{Bel}_{B} \oplus (\mathrm{Bel}_{B}^{\downarrow})_{\{B,\bar{B}\}}$ reduces to Bel_{B}. Equation 4.22 states that in order to determine the degrees of belief of the children of A resulting from all the evidence bearing on nodes below A, it is sufficient to consider each child separately. This represents the mechanism for passing up belief values between different levels of hierarchy; it satisfies the intuitive argument to the effect that any evidence confirming a node should also provide evidence confirming its parent (Ref. 50).

Generalizing for Θ, we obtain:

$$(\mathrm{Bel}_{\Theta}^{\downarrow})_{\ell_\Theta} = \oplus\{\mathrm{Bel}_B \oplus (\mathrm{Bel}_B^{\downarrow})_{\{B,\overline{B}\}} | B \in \ell_\Theta\}. \quad (4.23)$$

The partition becomes ℓ_Θ instead of $\ell_\Theta \cup \{\overline{\Theta}\}$, since $\ell_\Theta \cup \{\overline{\Theta}\} \subset \ell_\Theta$.

A statement even stronger than statement c. above is: if A is a non terminal element of $\mathfrak{R}$, then the partition $\ell_A \cup \{\overline{A}\}$ discerns the interaction relevant to itself among $\mathrm{Bel}_A^{\downarrow}$, Bel_A and $\mathrm{Bel}_A^{\Diamond}$. Therefore, exploiting (4.21) for $A \in \mathfrak{R}$ and the last statement, we obtain:

$$(\mathrm{Bel}_{\Theta}^{\downarrow})_{\ell_A \cup \{\overline{A}\}} = \{(\mathrm{Bel}_A^{\downarrow})_{\ell_A \cup \{\overline{A}\}} \oplus (\mathrm{Bel}_A)_{\ell_A \cup \{\overline{A}\}} \oplus (\mathrm{Bel}_A^{\Diamond})_{\ell_A \cup \{\overline{A}\}}\}$$

or, equivalently:

$$(\mathrm{Bel}_{\Theta}^{\downarrow})_{\ell_A \cup \{\overline{A}\}} = \{(\mathrm{Bel}_A^{\downarrow})_{\ell_A \cup \{\overline{A}\}} \oplus \mathrm{Bel}_A \oplus (\mathrm{Bel}_A^{\Diamond})_{\{A,\overline{A}\}}\} \quad (4.24)$$

since Bel_A is carried by $\ell_A \cup \{\overline{A}\}$ and $(\mathrm{Bel}_A^{\Diamond})_{\ell_A \cup \{\overline{A}\}} = (\mathrm{Bel}_A^{\Diamond})_{\{A,\overline{A}\}}$ from statement b. Equation 4.24 states that the evidence from above A and down other branches affects the degree of belief of the children of A only if the degree of belief is for or against A itself. This is the mechanism for passing, confirming and disconfirming evidence to the lower levels of the hierarchy.

4.5.3 The Shafer-Logan Algorithm

The algorithm proposed by Shafer & Logan (Ref. 3) is based on the concepts described in the previous subsection and Appendix C. The combination of hierarchical evidence is accomplished in four stages.

At stage 0, evidence is received for a specific node in the form of a simple support function or dichotomous function. This evidence is first combined with the existing belief value associated with the node using Dempster's rule of combination. In this case the combination is easy to calculate, since the belief functions being combined are simple support functions or dichotomous functions focused on the same node. The result of this simple combination is a dichotomous belief function which is then propagated through the hierarchical tree using the Shafer-Logan algorithm stages 1 to 3 below.

At stage 1, the dichotomous belief functions attached to terminal nodes are combined to find degrees of belief for and against their parents; the same is done to the parents' parents and so on, until there is a dichotomous belief function attached to each child of A to obtain the values of $(\mathrm{Bel}_{A}^{\downarrow})_{\ell_A \cup \{\overline{A}\}}$. This is performed by (4.22). This stage calculates degrees of belief by moving up the tree.

At stage 2, we obtain the values of each (direct) child of $\Theta : (\mathrm{Bel}_{\Theta}^{\downarrow})_{\ell_\Theta}$, using (4.23).

At the last stage, the degree of belief of each node is reevaluated to take into account the influence of other nodes using (4.24); this process calculates the degrees of belief by moving back down the tree. Figure 12 illustrates the algorithm's flow chart.

The implementation of the algorithm is not straightforward. The combination of dichotomous belief functions is performed using the commonality functions of (4.20). The formulas used to implement the algorithm are reproduced in Appendix C (Section C.1).

The amount of arithmetic involving a particular node depends linearly on the number of daughters of the node. Furthermore, the computational complexity of the algorithm is linear in the number of nodes in the tree. This is the case because the belief functions being combined are simple support functions focused on nodes or their complements. However, if we were to combine belief functions carried by the field of subsets generated by the children of a node, more precisely by a partition $\ell_A \cup \{\overline{A}\}$ (such as suggested by Shafer, Ref. 49), then the amount of arithmetic would become exponential in the sib size but remain proportional to the number of sibs.

Note that a special case occurs when the belief functions are Bayesian, that is, when the belief function Bel_A carried by $\ell_A \cup \{\overline{A}\}$ satisfies:

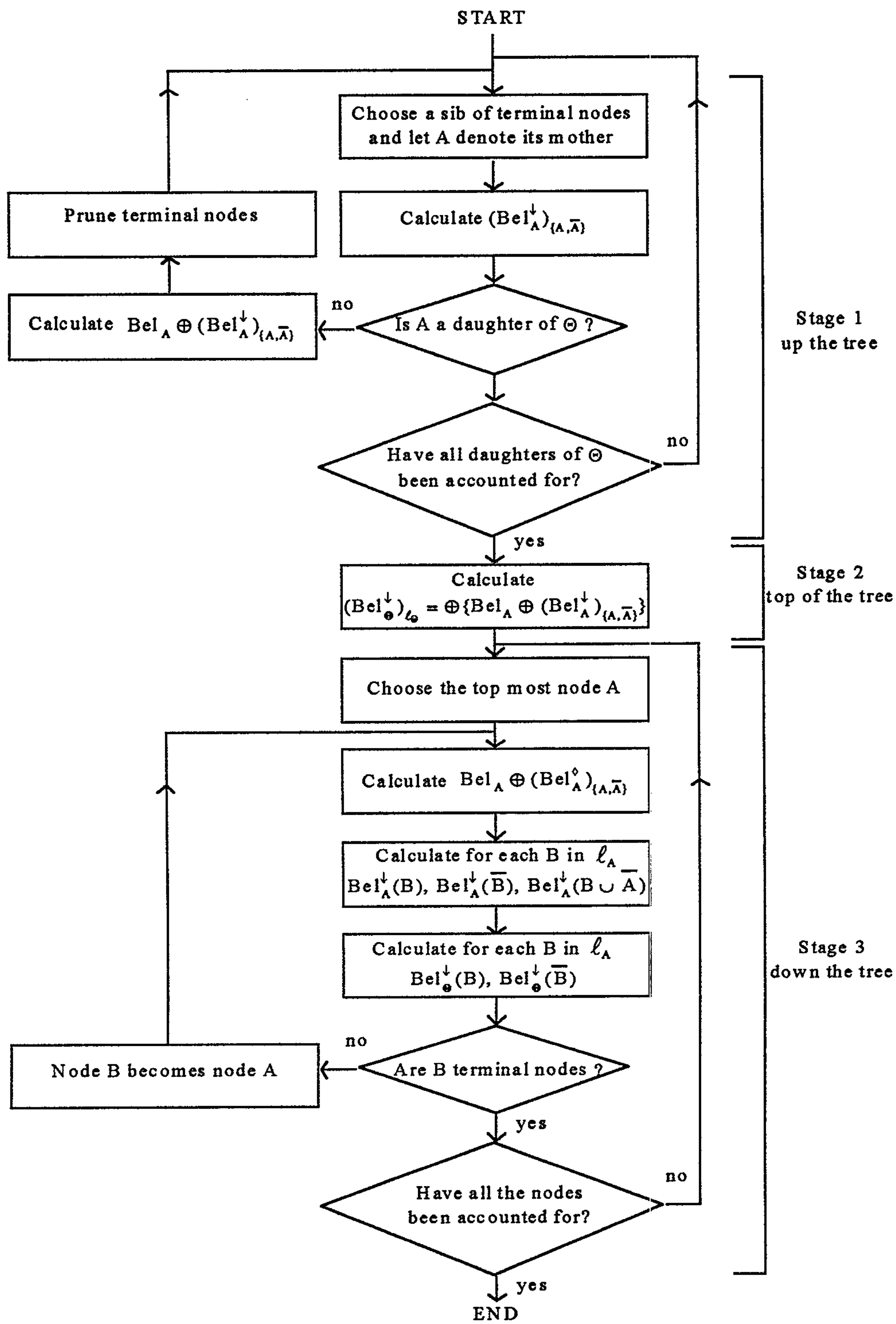


FIGURE 12 - Flowchart for the Shafer-Logan Algorithm

$$\mathrm{Bel}_A(B|A) + \mathrm{Bel}_A(\overline{B}|A) = 1$$

and

$$\mathrm{Bel}_A(B|\overline{A}) + \mathrm{Bel}_A(\overline{B}|\overline{A}) = 1$$

for every element B of the field $(\ell_A \cup \{\overline{A}\})^*$; the arithmetic is then linear in the sib size. This case is specifically the one devised by J. Pearl and described in Subsection 3.6.2.

The Dempster combination rule for combining belief functions, as described in Subsection 4.2.1, was implemented on an HP workstation using the C++ language. The Shafer-Logan algorithm for combining simple support functions focused on a subset of Θ or its complement was also implemented. Two examples are provided in Appendices. The first illustration (Appendix C, Section C.2) shows that if the belief functions are Bayesian, then Dempster's combination rule gives similar results to those obtained by Pearl's algorithm. This example uses the strict hierarchical tree illustrated in Figure 9 of Subsection 3.6.3. The *a priori* probabilities are indicated for each set of interest. For example,

$$m(\{B\}) = m(\{C, D\}) = P(B|E_1) = 0.1538.$$

Here, the Shafer-Logan algorim can m(2Unot be applied since the *a priori* probabilities are not dichotomous in nature.

The second example (Appendix C, Section C.3) shows the propagation effect of combining dichotomous belief functions using the Shafer-Logan algorithm. This example, in which 6 sets of evidence are combined, illustrates the propagation effect of the Shafer & Logan algorithm. The same strict hierarchical tree as above is used. The example is composed of 6 steps, at which additional evidence is received for a specific node in the form of a simple support function or dichotomous function, and then combined. The new belief (Bel) and plausibility (Pl) values are calculated for each node. Figures 13 to 18 show the results of the Shafer-Logan algorithm after adding evidence from step 1 to 6 respectively.

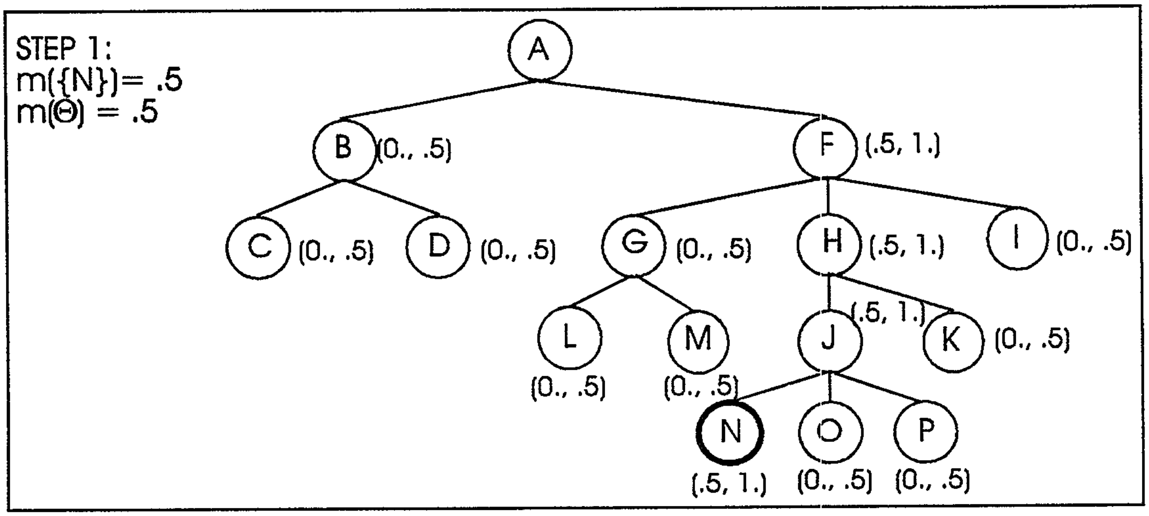


FIGURE 13 - Results From the Shafer-Logan Algorithm - Step 1

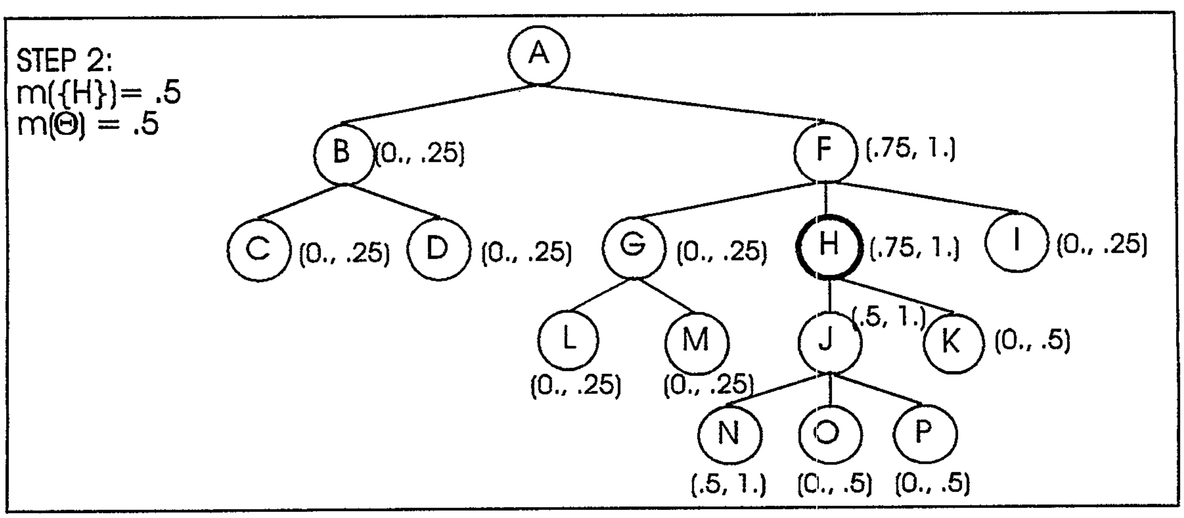


FIGURE 14 - Results From the Shafer-Logan Algorithm - Step 2

A plausibility of 1 for {N} implies that so far no evidence can refute {N} whereas a plausibility of .5 for {L} (for example) is obtained through evidence $m(\Theta) = .5$.

The evidence $m(\{H\}) = .5$ does not influence the children of {H} (in terms of belief and plausibility) but influences the belief of the father of {H} and the plausibility of the other nodes.

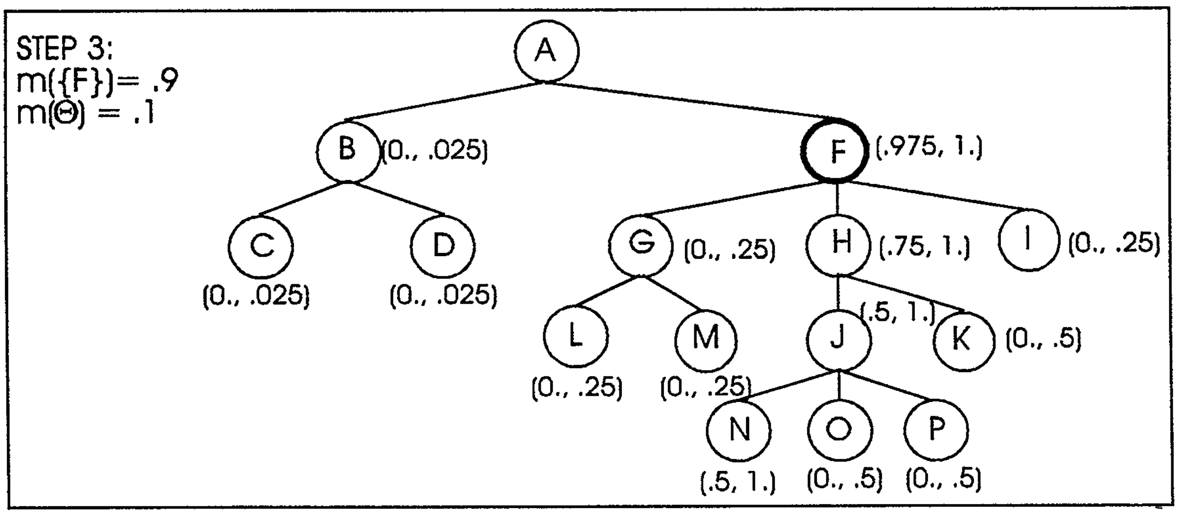


FIGURE 15 - Results From the Shafer-Logan Algorithm - Step 3

In a similar fashion, a strong evidence for {F} does not affect the belief of {B}; however, it diminishes the plausibility of {B}.

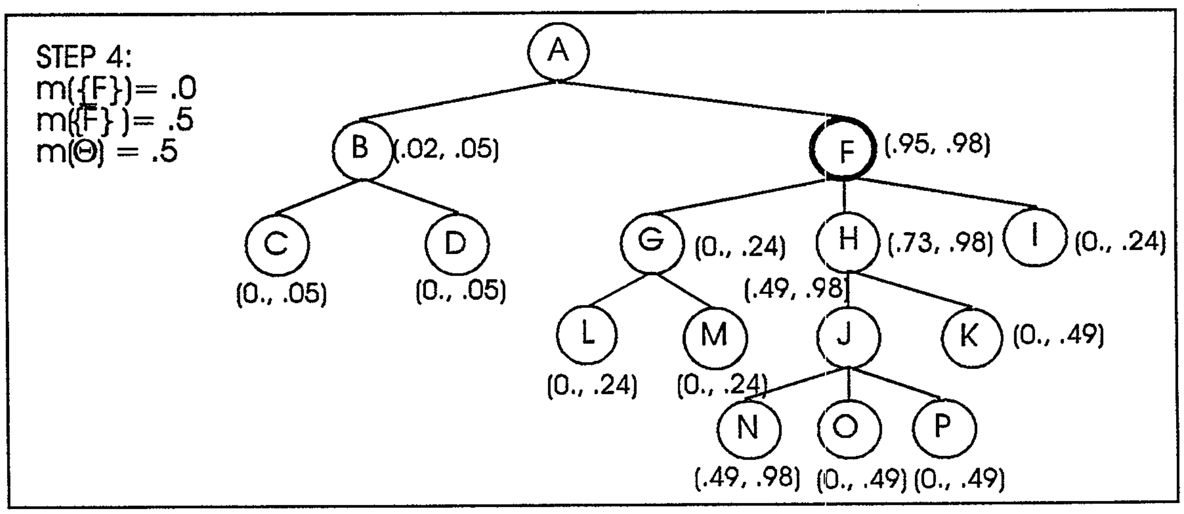


FIGURE 16 - Results From the Shafer-Logan Algorithm - Step 4

As predicted, the evidence against {F} $(m(\{\overline{F}\}) = .5)$ affects the children of {F}.

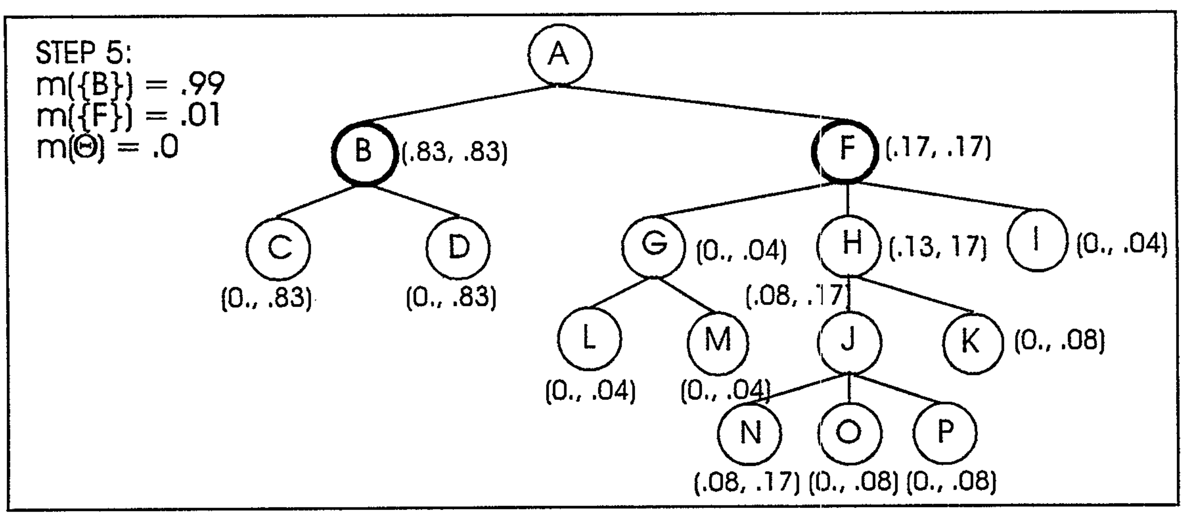


FIGURE 17 - Results From the Shafer-Logan Algorithm - Step 5

The contradictory evidences greatly modifies the belief and plausibility of {B} and {F}, when combining the two dichotomous belief functions focused on the same nodes B and F (m({B}) = .99, m({F}) = .01 with m({B}) = .02, m({F}) = .95, m(Θ) = .03). Dempster's rule calculates a small normalizing constant (K = .06), indicating conflict. It is interesting to note that this evidence is dichotomous since {B} = $\{\bar{F}\}$.

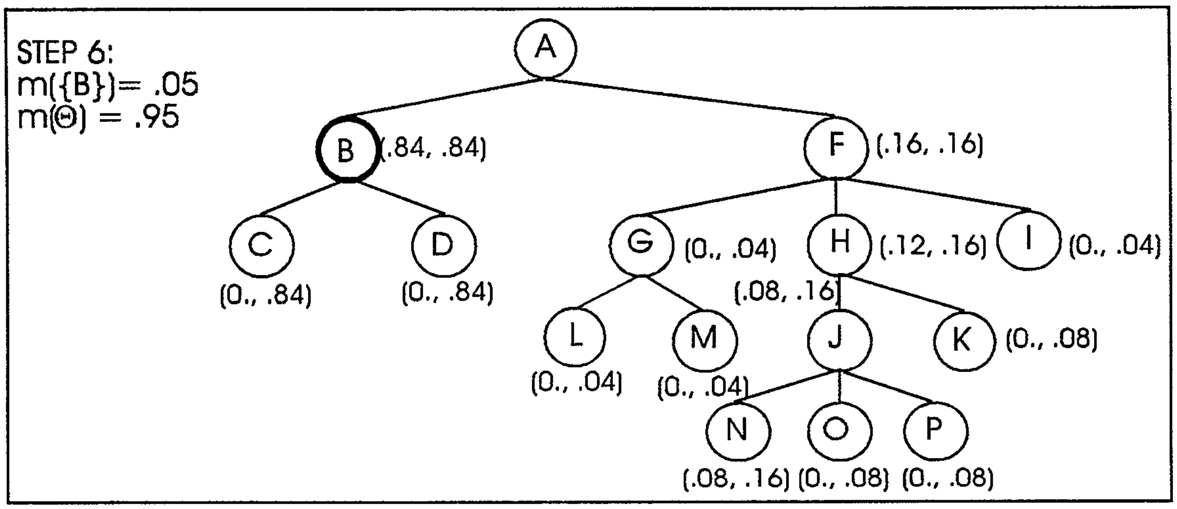


FIGURE 18 - Results From the Shafer-Logan Algorithm - Step 6

The small evidence for {B} slightly modifies the belief and plausibility of {B}; the big uncertainty is distributed amongst the other nodes.

5.0 STATISTICAL DECISION MAKING

An important element to take into consideration in the design of an identity declaration fusion function is the decision making process required to select the identity declaration which best supports the combined declarations (Barnett, Ref. 25). Statistical decision making is necessary since, after fusion, the resulting hierarchical structure may contain many identity declarations with a non-null confidence value. Because decisions are subjective, the decision maker will undoubtedly depend on his or her own judgment as well as information collected from various sources. The fusion function should still suggest to the decision maker the "best" candidate or candidates according to predetermined decision rules. A number of factors are involved in the choice of the decision rules to be used (Nahim & Pokoski, Ref. 18) :

- rule complexity (if the computational time required to make a decision is too high, the process becomes useless),
- confidence level (decision rules are made on a probabilistic basis),
- fusion technique used to combine uncertain information (for example the Dempster-Shafer approach produces belief and plausibility values instead of single probability values),
- type of application (the application environment can dictate decision rules).

Before analyzing the specific needs of an identity declaration fusion function, various approaches will be discussed concerning decision techniques in the face of knowledge combined by the Dempster-Shafer theory. Then, decision making will be studied pertaining to information structured in a hierarchical manner, in order to provide a decision making approach to the identity declaration fusion function.

5.1 Statistical Decision Making Based on the Dempster-Shafer Representation

The Dempster-Shafer theory has been applied in various contexts. However, no general method is acknowledged for classification (decision making) based on basic probability assignments (Liu & Yang, Ref. 51). It is important to note that the Dempster-

Shafer belief calculus provides two measures for decision making: the belief (Bel) and plausibility (Pl) measures. Different approaches have been studied based on Bel and/or Pl; these will be briefly discussed.

Selzer & Gutfinger (Ref. 52) have proposed a method based on the belief measure Bel accompanied by heuristic rules to choose the best alternative among many possible alternatives:

- the best alternative must have the maximum basic probability number,
- the difference in basic probability numbers between any two alternatives must be above a specified threshold.

For their classification problem, Liu & Yang (Ref. 51) also proposed a method based on the belief measure Bel, but added more rules for selecting the best alternative:

- the best alternative must have the maximum basic probability number,
- the difference between the basic probability number of the best alternative and the other alternatives should be larger than a threshold,
- the basic probability number for uncertainty $m(\Theta)$, should be less than a certain threshold,
- the basic probability number for the best alternative should be larger than the basic probability number for uncertainty $m(\Theta)$.

If the constraints are not met, the method does not propose a best alternative.

In the Paramax (Ref. 7) study, a modified Dempster-Shafer approach was used to fuse primarily attribute information in order to obtain identity classification. The decision rule proposed was based on the maximum basic probability number of an alternative chosen among certain important alternatives which were of tactical or strategic interest. Therefore, not all alternatives were candidates for being the best alternative; the subset of candidates was selected according to the scenario and mission at hand.

Voorbraak (Ref. 46) suggested the use of the belief interval [Bel({a}), Pl({a})] for decision making. As there is no unique way of ordering the belief intervals with respect to their degree of certainty, he proposed four orderings induced by the following rules:

1. The minimal ordering $\leq_{min}$ is defined by $[x,y] \leq_{min} [x',y']$ iff $y \leq x'$
2. The ordering by average $\leq_{av}$ is defined by $[x,y] \leq_{av} [x',y']$ iff $(x+y)/2 \leq (x'+y')/2$
3. The belief ordering $\leq_{Bel}$ is defined by $[x,y] \leq_{Bel} [x',y']$ iff $x \leq x'$
4. The plausibility ordering $\leq_{Pl}$ is defined by $[x,y] \leq_{Pl} [x',y']$ iff $y \leq y'$

The choice for minimal ordering corresponds to a rather cautious approach to the ordering of elements with respect to their certainty, whereas the choice for the ordering by average appears rather audacious in nature. The belief ordering simply corresponds to the maximum basic probability number. The plausibility ordering can play an important role since it indicates the extent to which the belief may vary. For example,

[Bel({a}), Pl({a})] = [0.5, 0.6] indicates that prob({a}) can vary between 0.5 and 0.6, but
[Bel({b}), Pl({b})] = [0.4, 0.8] indicates that prob({b}) can vary between 0.4 and 0.8.

Therefore, even if Bel({a}) > Bel({b}), the evidence suggests that prob({a}) cannot be higher that 0.6 whereas prob({b}) could be as high as 0.8.

Voorbraak (Ref. 46) does not favor one ordering over the other but mentions that the plausibility ordering can play a dominant role in the decision making process. Therefore in many situations, if using the belief ordering, he suggests that the plausibility of the best alternative be higher than the plausibility of all the other alternatives.

Barnett (Ref. 25) also adheres to the idea that the measures Bel and Pl should be used to assist decision making. However, he concentrated his own efforts on problems where most elements of Θ have basic probability numbers equal to 0. This occurs when a large number of evidence sources are not available. In such a case, he argues that Pl generally provides some discrimination even when the evidence is sparse. Therefore, he suggests that Pl is a more robust guide to decision making than is Bel. This concept was applied by Altoft (Ref. 53) to a classification problem in which his main decision criterion was to choose the alternative with the highest plausibility value, reserving the belief value for tie breaking.

5.2 Statistical Decision Making Based on the Dempster-Shafer Representation using a Hierarchical Structure

When dealing with a hierarchical structure, the decision making techniques of the previous section cannot be directly applied because the belief of a parent will always be equal to, or higher than, the belief of each of his children. Therefore, one cannot simply rely on the maximum value of belief. For example, in Figure 18, the belief of {F} is always higher than the belief of {G}, {H} or {I}. Similarly, the belief of {H} is higher than that of {J} or {K}. Also, the plausibility of a parent will always be equal to, or higher than, the plausibility of each of his children. An exception to this rule would be if the decision maker were interested only in the leaf nodes (the elements of the frame of discernment Θ), in which case the hierarchical structure would be superfluous.

Furthermore, an important aspect that should not be forgotten in military applications, as already mentioned by Liu & Yang (Ref. 51) and Paramax (Ref. 7), is the fact that decision making is scenario and mission dependent.

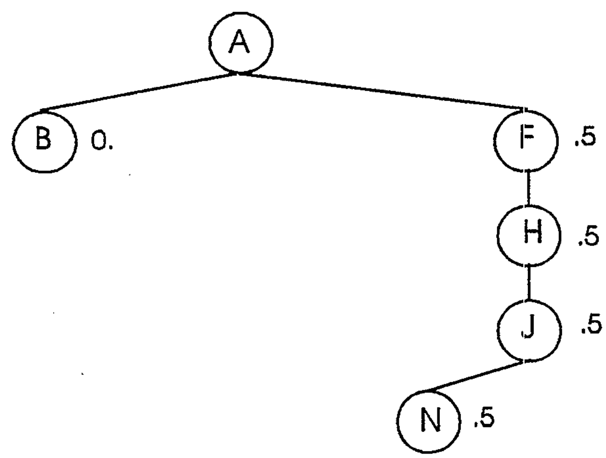
What we propose, therefore, is a semi-automated approach based on the belief and plausibility values. It is called semi-automated because threshold values will be applied to the belief and plausibility measures. However, the final decision will be taken by the decision maker, because he/she remains an important part of the process and because the choice of the final identity is typically scenario and mission dependent. The decision making approach proposed is as follows:

- select all alternatives with a plausibility value greater than a certain threshold T_{Pl} ;
- plot the chosen alternatives according to hierarchical structure and indicate for each node its belief value;
- add to the graph all the nodes directly below Θ with their belief values.

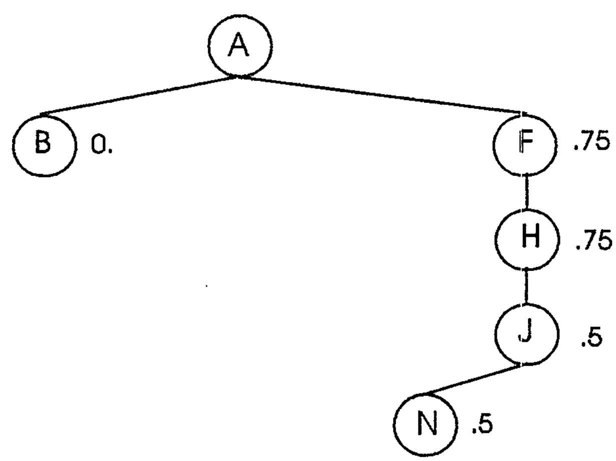
Based on the graph constructed by this decision approach, the decision maker can select the best alternative according to the highest belief value, or the hierarchical level of interest or any other criteria. The plausibility value is not included in the graph as it is not deemed to be essential to the decision making process. To illustrate the approach, the decision technique will be applied to the example of Section C.3 (Appendix C).

The plausibility threshold T_{Pl} is chosen equal to .6. At each step, the decision technique creates a graph from which the decision maker can select the best alternative according to his/her needs. If no nodes other than the direct children of Θ appear in the graph, then no alternative has a plausibility value higher than the threshold T_{Pl} and no decision can be taken. The results are as follows:

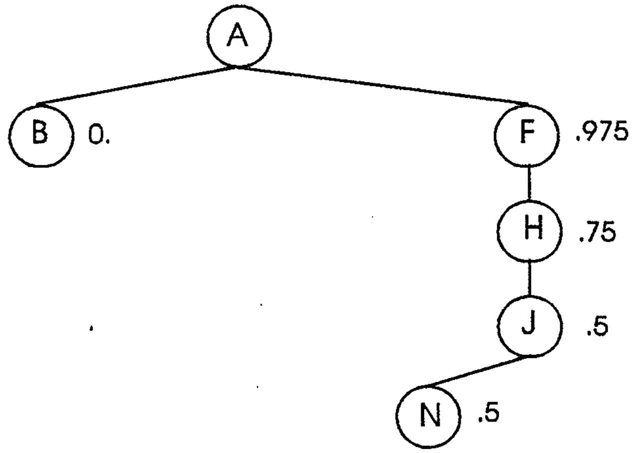
STEP 1:



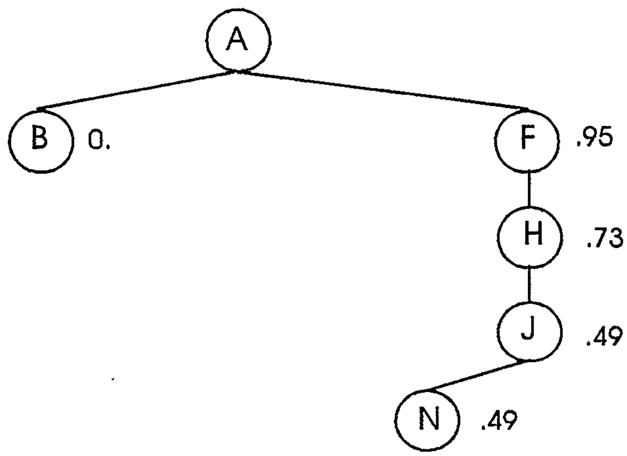
STEP 2:



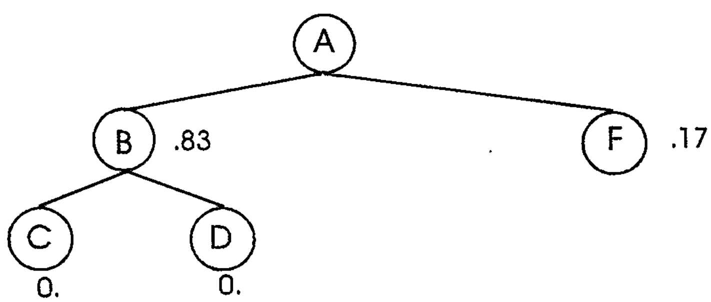
STEP 3:



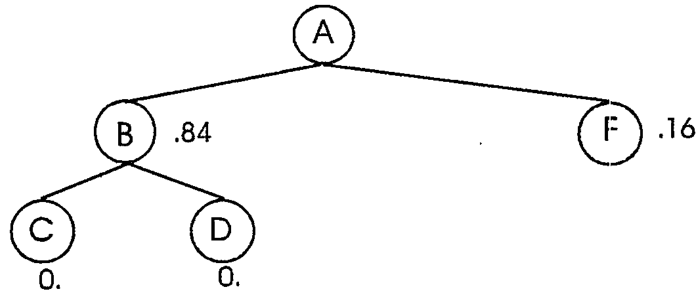
STEP 4:



STEP 5:



STEP 6:



As indicated in Chapter 4, the contradictory evidence at step 5 greatly modifies the belief of {B} and {F}. The decision technique reproduces the effect of this contradictory evidence.

6.0 IDENTITY DECLARATION FUSION FUNCTION

We are now able to provide an identity declaration fusion function based on the various concepts studied in the previous chapters.

The first section describes the identity declaration fusion function and the second section provides an example using the identity declaration fusion function in which identity declarations will be fused and the decision making process applied.

6.1 Description of the Identity Declaration Fusion Function

Within the framework of our identity information fusion study, various hypotheses were made and delineated in the previous chapters; the two hypotheses which bear the most impact on the choice of the fusion approach are as follows:

1. The evidence provided by the various information sources are independent according to Shafer's definition.
2. Probabilistic information is only available for some of the events associated with Θ, such that the likelihood matrix is not fully specified.

These hypotheses suggest that the Dempster-Shafer theory of evidence is an appropriate technique to fuse uncertain information. Because we have chosen to represent identity declarations in a hierarchical manner, the algorithm proposed by Shafer and Logan is appealing, both in terms of the information structure and computational requirements.

An interesting issue which distinguishes this study from other studies on identity fusion is the fact that 2 different frames of discernment are introduced to estimate the identity of the observed objects. The first frame of discernment is the hierarchical tree of surface and air classifications, as shown in Figures 5a and 5b, and the second one deals with the threat categories as described in Figure 19. The purpose of this approach is to circumvent the problem whereby the threat category of a detected object is directly inferred from its identity. For example, in Figure 19 the Exocet missile is automatically assumed friendly. During the Falklands war, however, the Exocet missile was definitely not considered friendly to the British Navy. By eliminating false automated inferences, this approach allows more freedom in the decision making.

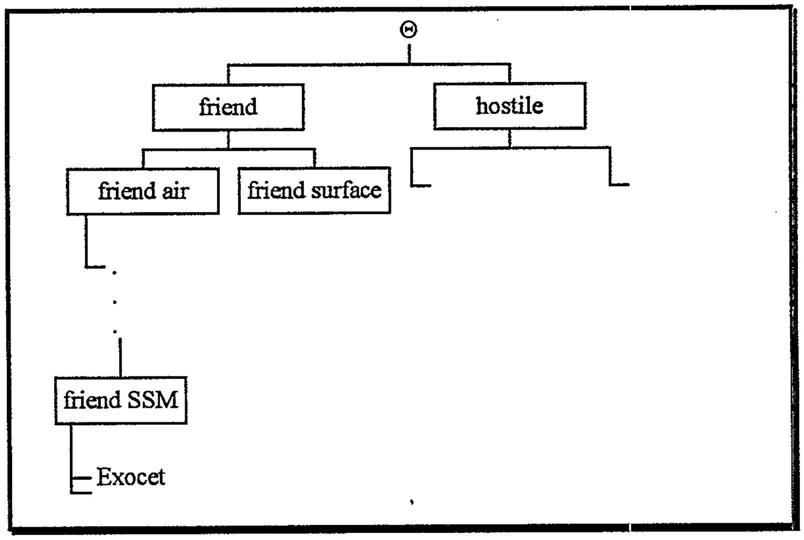


FIGURE 19 - Example of Frame of Discernment Where Threat Category is Directly Inferred From Identity

To accommodate a hierarchical structure, Figure 6 must be modified, as shown in Figure 20, to include Θ_2. The "pending" subdivision was eliminated since an uncertainty value for this element would not be available. Also, the "suspect" and "assumed friend" subdivisions had to be eliminated because they do not form a set of mutually exclusive events with the "hostile" and "friend" subdivisions (according to the definition of a frame of discernment). In our opinion, the two frames of discernment Θ_1 and Θ_2 of Figure 20 could encompass most of the identity declarations within the naval environment.

A general fusion approach is now proposed based on the concepts discussed in the previous chapters.

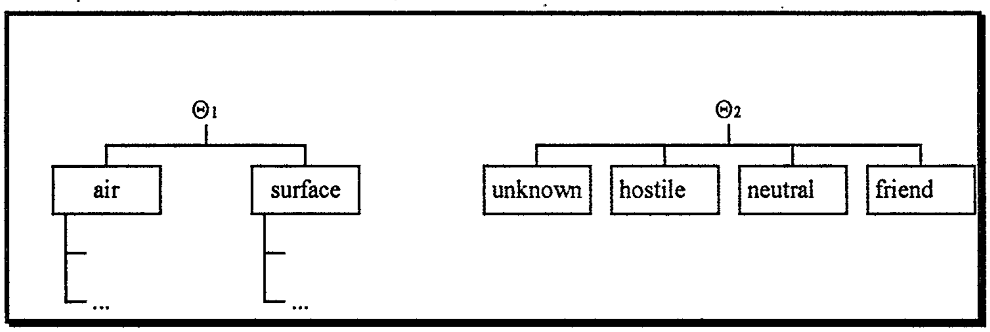


FIGURE 20 - The Two Frames of Discernment Used in the Identity Declaration Fusion Function

The most basic concept to which the study adheres is the fact that information sources are self contained and that each one represents a local decision node capable of identifying a detected object (Section 2.2). The aim is, therefore, to combine local decisions in the hope of obtaining the correct identity with a high probability. Dasarathy (Ref. 54) calls this type of fusion "Decision In-Decision Out Fusion" because both the input and output are decisions. An appropriate architecture to delineate this concept is sensor level architecture (Subsection 2.3.2.1). Figure 21 reproduces a simplified version of sensor level architecture from Figure 7.

In the case of multiple objects in the detection environment, the association process matches the received identity declaration, originating from an information source, with one of the observed objects. Identity declaration is one of many components that characterize a detected object; these components form the state vector of the object and are called a *track*. Each detected object is typified by its track. In the event that an information source provides identity declaration for an existing track, the identity declaration fusion function becomes necessary to combine these declarations. Consequently, the fusion process is applied to each track based on the frame of discernment (Θ_1, Θ_2 or both) according to the type of identity declarations to combine. For each track, the frames of discernment are identical; however, the belief and plausibility values vary according to the weight of the evidence received.

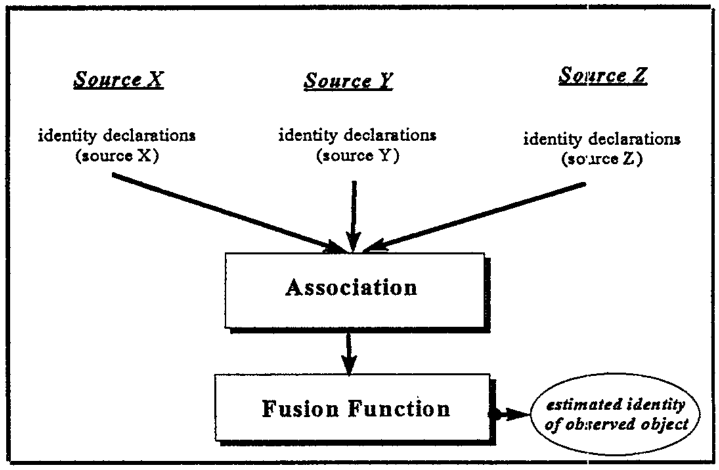


FIGURE 21 - Identity Declaration Fusion Process - Sensor Level Architecture

It is noteworthy that the order in which bodies of evidence are received is inconsequential since the orthogonal sum of the Dempster-Shafer technique is commutative and associative. Therefore, if two information sources simultaneously provide identity declarations on the same observed object, the combination of these evidences with the existing information will result in identical belief and plausibility values, whether one evidence is added before the other.

If we assume that the association function performs perfectly, the fusion function can be illustrated by the flowchart of Figure 22. Its major processes are outlined below.

1. As described in Subsection 2.3.2.2, it was assumed that all sources capable of providing identity declarations will do so by attaching to each declaration a quantitative measure of uncertainty. This measure corresponds to the probability that the identity declaration and detected object are matched or, equivalently, to the probability that declaration i from source s is true:

$$\begin{aligned} C_{s,i} &= \text{P(declaration } i \text{ from source } s \text{ matches detected object)} \\ &= \text{P(declaration } i \text{ from source } s \text{ is true).} \end{aligned}$$

   In the case of non-sensor information sources, the matching coefficient $C_{s,i}$ simply typifies a subjective confidence appraisal of the declaration.

2. It is further assumed that information sources provide single identity declarations as opposed to multiple declarations:

example of single identity declaration:	military fixed wing, $C_{s,i} = 0.7$
example of multiple identity declaration:	military fixed wing, $C_{s,i} = 0.4$
	and civil fixed wing, $C_{s,i} = 0.1$.

Thus, it is simple to transform the probability of a true declaration into a dichotomous or simple support function. Effectively, if the single declaration is military fixed wing with $C_{s,i} = 0.7$, then we obtain the following simple support function:

$$m(\text{military fixed wing}) = 0.7$$
$$m(\Theta) = 0.3.$$

If the single declaration is military fixed wing with $C_{s,i} = 0.7$, not military fixed wing with $C_{s,i} = 0.1$, then we obtain the following dichotomous belief function:

$$m(\text{military fixed wing}) = 0.7$$
$$m(\overline{\text{military fixed wing}}) = 0.1$$
$$m(\Theta) = 0.2.$$

3. The dichotomous or simple support function is then combined with the belief value, or more precisely the dichotomous belief function of the same focal element within the hierarchical tree; this is accomplished using Dempster's combination rule as explained at stage 0 of Subsection 4.5.3. However, as shown in Subsection 4.3.2,

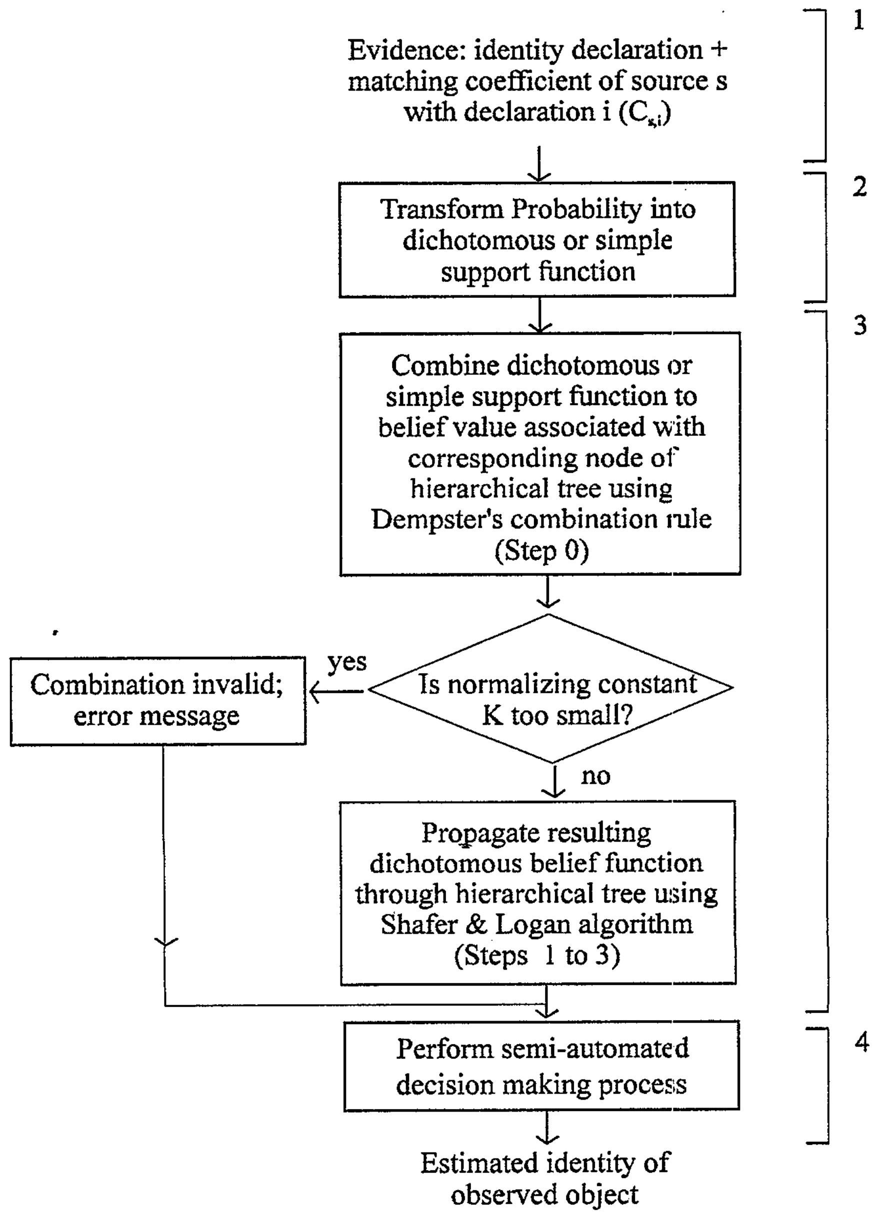


FIGURE 22 - Flowchart of Identity Declaration Fusion Function

the normalization inadequacies of Dempster's combination rule can pose a serious problem. Unfortunately, Yager's degree of informativeness cannot be applied since we are dealing with simple support or dichotomous functions. We could, however, easily determine the degree of conflict when combining simple support or dichotomous functions; this can be accomplished at stage 0, as described in

Subsection 4.5.3. Therefore, if the normalizing constant K is greater than an appropriate threshold value, the information is said to be in relative agreement and Dempster's combination rule can be applied. If the normalizing constant is smaller than the threshold, the bodies of evidence are conflicting, suggesting that one of the assessments is unreliable, signaling a potential information source problem (Abdulghafour & Abidi, Ref. 55). It is, therefore, suggested that the combination be suspended pending verification of the information sources.

If the combination is valid, the result of the combination is propagated through the hierarchical tree using the Shafer-Logan algorithm.

4. The last phase of the fusion function performs the semi-automated decision making process in the sense that the decision maker must select the best alternative among a limited subset of likely identity declarations.

6.2 Example of the Fusion Function Applied to the Problem of Identity Declarations

The following example illustrates the use of the Shafer-Logan algorithm for the specific problem of combining identity declarations using two frames of discernment. The first frame of discernment Θ_1 is detailed in Figure 23. The hierarchy has been simplified from that of Figures 5a and 5b. The names of the leaf nodes (for example MiG-19) were taken from Refs. 56-57 and are both friendly and hostile elements. The second frame of discernment Θ_2 is exactly as described in Figure 20; in other words the leaf nodes are those of Figure 20.

The scenario is as follows: the commander of a Canadian Patrol Frigate-type ship receives a series of identity declarations concerning one object; he/she must determine the identity of the object and take action.

As in the previous examples, the belief and plausibility values at each node of the two hierarchical trees are zero. The evidences received from various information sources are given in Table III (assuming that the order of the received evidences is unimportant):

TABLE III

Evidences received from various sources

Frame of discernment Θ_1	Frame of discernment Θ_2
$m(\{\text{fighter}\}) = 0.4$	$m(\{\text{unknown}\}) = 0.2$
$m(\Theta) = 0.6$	$m(\Theta) = 0.8$
$m(\{\overline{\text{carrier}}\}) = 0.7$	$m(\{\overline{\text{friend}}\}) = 0.6$
$m(\Theta) = 0.3$	$m(\Theta) = 0.4$
$m(\{\text{fixed-wing}\}) = 0.3$	$m(\{\text{hostile}\}) = 0.7$
$m(\Theta) = 0.7$	$m(\Theta) = 0.3$
$m(\{\overline{\text{non-combatant}}\}) = 0.8$	$m(\{\text{neutral}\}) = 0.3$
$m(\Theta) = 0.2$	$m(\Theta) = 0.7$
$m(\{\text{air}\}) = 0.5$	$m(\{\text{hostile}\}) = 0.8$
$m(\Theta) = 0.5$	$m(\Theta) = 0.2$
$m(\{\overline{\text{air}}\}) = 0.1$	
$m(\Theta) = 0.9$	
$m(\{\text{MiG-25}\}) = 0.1$	
$m(\Theta) = 0.9$	
$m(\{\text{helicopter}\}) = 0.5$	
$m(\Theta) = 0.5$	
$m(\{\text{MiG-19}\}) = 0.6$	
$m(\Theta) = 0.4$	
$m(\{\overline{\text{missile}}\}) = 0.7$	
$m(\Theta) = 0.3$	
$m(\{\text{MiG-25}\}) = 0.4$	
$m(\Theta) = 0.6$	

The evidences are the only ones received concerning the object; a decision could be taken after each evidence if the plausibility of at least one node is greater than T_{Pl}. However, we have chosen to combine all the evidences before the decision making process. In this example, T_{Pl} will be set to .5. Evidences $m(\{\text{air}\}) = .5$ and $m(\{\overline{\text{air}}\}) = .1$ are contradictory but not enough for the Dempster's combination rule to produce irregular results. Figures 24 and 25 show the belief and plausibility measures for each node after the

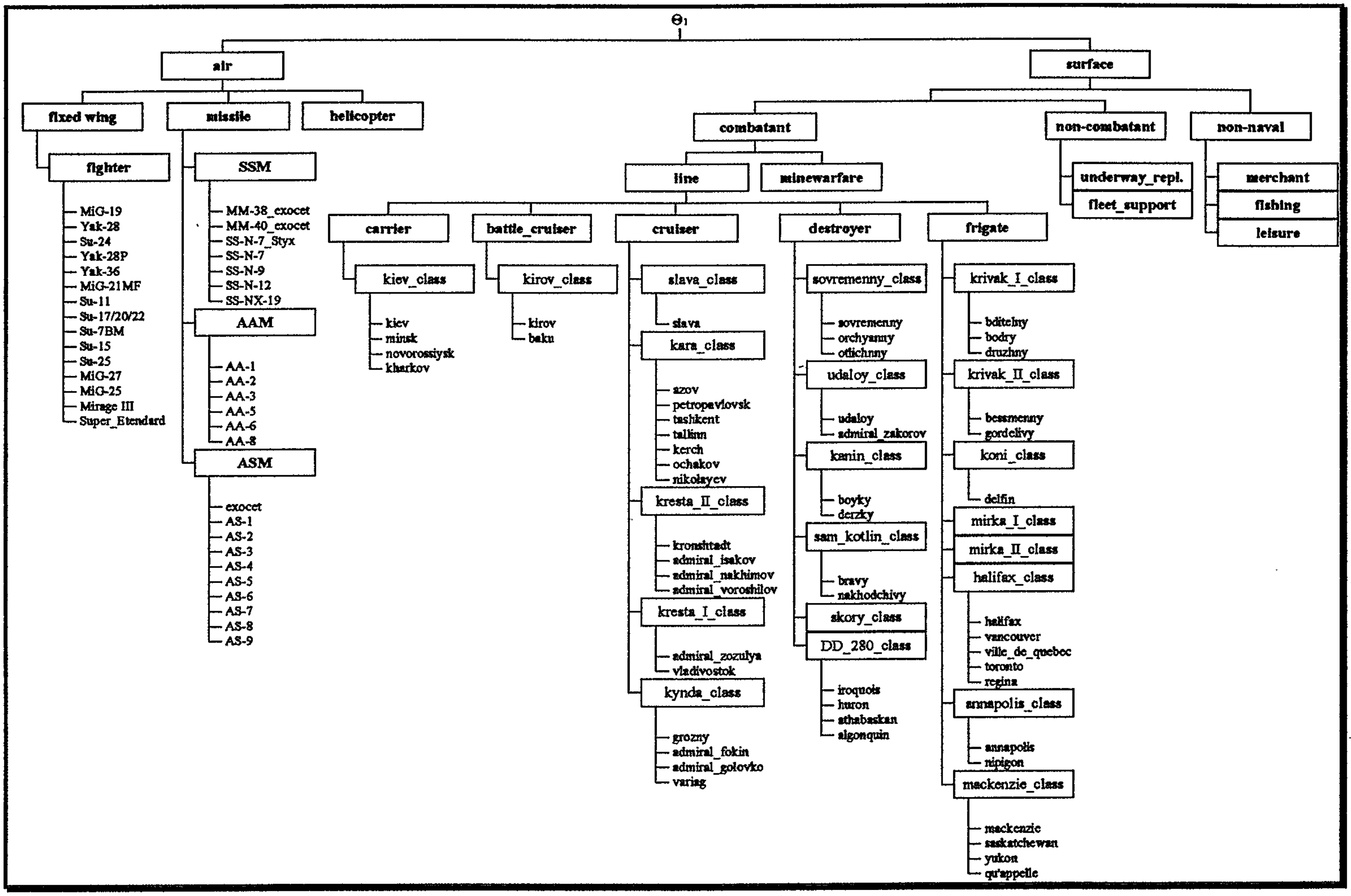


FIGURE 23 - Frame of Discernment Θ_1 Used in Example

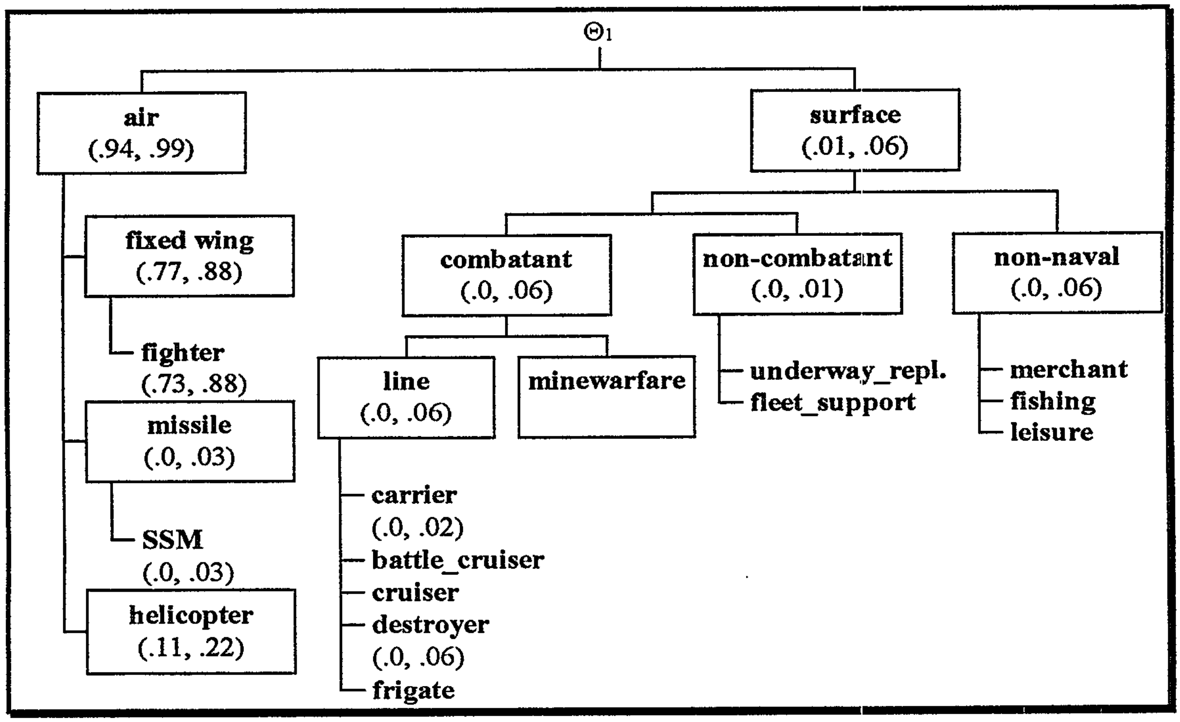


FIGURE 24 - Results From Combining Evidences Using Θ_1

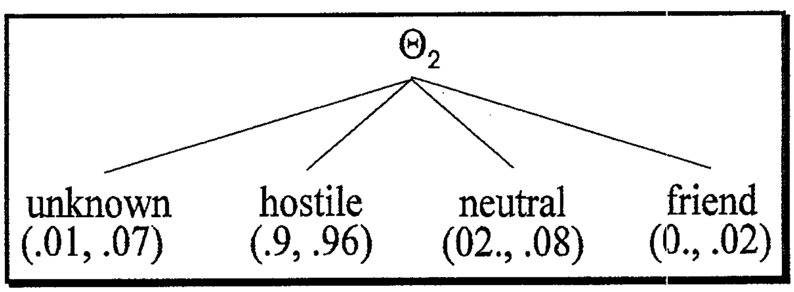


FIGURE 25 - Results From Combining Evidences Using Θ_2

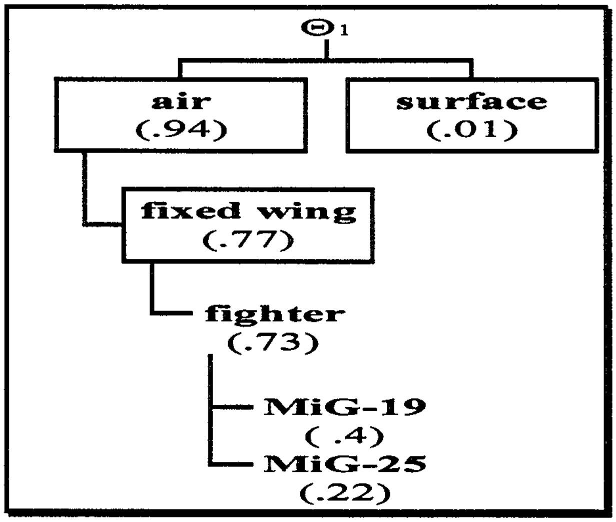


FIGURE 26 - Graph Representing Best Alternatives Using Θ_1

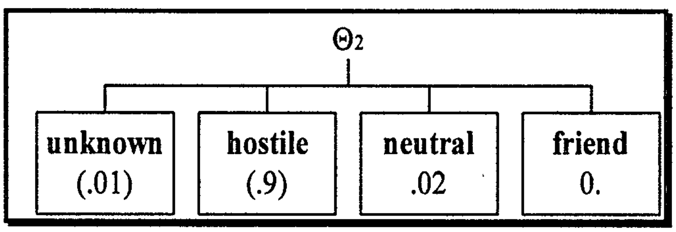


FIGURE 27 - Graph Representing Best Alternatives Using Θ_2

combination of evidences for frames of discernment Θ_1 and Θ_2, respectively. Figures 26 and 27 are the graphs available to the decision maker. It seems that the object is airborne and hostile,and there is a fairly good chance that it is a fixed wing fighter. According to the mission, the decision maker will choose the best alternative and take appropriate action if necessary.

7.0 CONCLUSION

This document is concerned with the use of the Dempster-Shafer theory of evidence for the fusion of identity declarations within a naval environment. It proposes to hierarchically structure the identity declarations according to NATO's STANAG 4420 charts, which provide a better base for achieving interoperability in information exchange between nations than uncontrolled alternatives.

The Bayesian approach is also investigated but is found to suffer from major deficiencies in a hierarchical context, when fully specified likelihoods are not available. Other problems associated with this approach are the coding of ignorance, and the strict requirements on the belief of a hypothesis and its negation.

One drawback of the Dempster-Shafer evidential theory is the long calculation time required by its high computational complexity. Due to the hierarchical nature of the evidence, an algorithm proposed by Shafer & Logan (1987) is implemented which reduces the calculations from exponential to linear time, proportional to the number of nodes in the tree.

A semi-automated decision making technique, based on belief and plausibility values, is then described for selecting alternatives which best support the combined identity declarations. The final decision will be taken by the decision maker, because he/she remains an important part of the process and because the choice of the final identity is typically scenario and mission dependent.

The use of the Dempster-Shafer theory of evidence in this document shows it to be a logical method of combining data from various sources to help the commander carry out his/her duties. However, the flexibility of the approach should not hide its shortcomings. For example, the normalization constant from Dempster's combination rule may give inaccurate results, and the independence requirements may sometimes be difficult to prove. Also, the final frame of discernment for threat categories (Θ_2), which had to be simplified due to the hierarchical constraints of the Dempster-Shafer technique, may not be sufficiently detailed for the needs of the commander. Lastly, because no general method for decision making from hierarchical evidence is acknowledged in the literature, simple heuristic methods, such as the one proposed in Chapter 5, are usually applied.

This document presents initial results of investigations on the use of the Dempster-Shafer approach in the naval environment. Nevertheless, the results show that the various concepts studied could be applicable to the domain of wide area fusion within the framework of a Communications, Command, Control and Intelligence (C^3I) system.

8.0 REFERENCES

1. Wilson, G. B. (1987). Some aspects of data fusion. *Advances in Command, Control & Communications Systems, IEE Computing Series*, **11**, 321-338.

2. Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan* Kaufmann Publishers, San Mateo, CA, 552 p.

3. Shafer, G. & R. Logan (1987). Implementing Dempster's rule for hierarchical evidence. *Artificial Intelligence*, **33**, 271-298.

4. Waltz, E. & J. Llinas (1990). *Multisensor Data Fusion.* Artech House Inc. New York. 464p.

5. Gibson, C. (1992). *The Management of Organic and Non-organic Information in a Maritime Environment: Analysis Requirements*. AUS-CAN-NZ-UK-US C&C Committee, 1-10.

6. Hall, D. L. (1992). *Mathematical Techniques in Multisensor Data Fusion.* Artech House Inc. New York, 301 p.

7. Paramax Systems Canada (1993). Investigations of attribute information fusion techniques for estimating target identity. Contract No. W7701-2-0400/01-XSK. Volumes 1 and 2.

8. Filippidis, A. & J. G. Schapel (1992). *Automatic Allocation of Identity.* Working Paper ITD/CSI-WP-92-4, Electronics Research Laboratory, Defence Science and Technology Organization, Australia.

9. Paramax Systems Canada (1992). Introduction to the Canadian Patrol Frigate Combat System.

10. Smith, C. R. & P. M. Goggans (1993). Radar target identification. *IEEE Antennas and Propagation Magazine*, **35(2)**, 27-38.

11. Foxwell, D., M. Hewish & G. Sundaram (1992). New naval electronic warfare: Integrating EW and C&C. *International Defense Review*, 7, 649-658.

12. Naval Forces Update. (1992). Improved EW on the horizon. *Jane's Defense Weekly*, 15 August 1992, 35.

13. Ferris, N. A., D. Gaucher, J. Kearney & R. Parker (1987). *Target Identification Expert Development*. RADC-TR-87-48. Rome Air Development Center, NY.

14. Dumas, J., and Sévigny, L. *Multisensor Image Fusion Project*, DREV M-3140/93, September 1993, UNCLASSIFIED.

15. Donker, J. C. (1991). Reasoning with uncertain and incomplete information in aerospace application. *AGARD Symposium on Machine Intelligence for Aerospace Electronic Systems*, Lisbon, **30**, 30.1-30.16.

16. Bégin, F., S. Kamoun & P. Valin (1994). On the implementation of AAW sensor fusion on the Canadian Patrol Frigate. *SPIE Conference, Vol. 2235, Signal and Data Processing of Small Targets*, 1-10.

17. Simard, M.-A., P. Valin & E. Shahbazian (1993). Fusion of ESM, Radar, IFF data and other attribute information for target identity estimation and a potential application to the Canadian Patrol Frigate. *66th NATO AGARD Avionics Panel Symposium on Challenge of Future EW System Design*, Ankara, Turkey, 18-21 October.

18. Nahim, P. & J. Pokoski (1980). NCTR plus sensor equals IFFN or can two plus two equal five? *IEEE Transactions on Aerospace & Electronic Systems*, **16**, 320-337.

19. Bogler, P. L. (1987). Shafer-Dempster reasoning with applications to multisensor target identification systems. *IEEE Transactions on Systems, Man and Cybernetics,* **SMC-17**, 968-977.

20. Buede, D. M., J. W. Martin & J. C. Sipos (1989). Comparison of Bayesian and Dempster-Shafer fusion. *Proceedings of the Data Fusion Symposium*, 81-90.

21. Hong, L. & A. Lynch (1993). Recursive temporal-spatial information fusion with applications to target identification. *IEEE Transactions on Aerospace and Electronic Systems*, **29**, 435-444.

22. STANAG 4420. Display Symbology and Colours for NATO Maritime Units. Edition 1. NATO UNCLASSIFIED.

23. Bhatnagar, R. K. & L. N. Kanal (1986). Handling uncertain information: A review of numeric and non-numeric methods. In *Uncertainty in Artificial Intelligence* (L. N. Kanal and J. F. Lemmer, eds). North-Holland, Amsterdam, 3-26.

25. Barnett, J. A. (1991). Calculating Dempster-Shafer plausibility. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**, 599-602

26. Billingsley, P. (1986). *Probability and Measure*, John Wiley & Sons.

27. Fagin, R. & J. Y. Halpern (1991). Uncertainty, belief, probability. *Computer Intelligence*, **7**, 160-173.

28. DeGroot, M. H. (1975). *Probability and Statistics*. Addison-Wesley Publishing Company, Reading, MA, 607 p.

29. Bacchus, F. (1990). Lp, a logic for representing and reasoning with statistical knowledge. *Computer Intelligence*, **6**, 209-231.

30. Grzymala-Busse, J. W. (1991). *Managing Uncertainty in Expert Systems*, Kluwer Academic Publishers, 224 p.

31. Henkind, S. J. & M. C. Harrison (1988). An analysis of four uncertainty calculi. *IEEE Transactions on Systems, Man and Cybernetics*, **18**, 700-714.

32. Rao, B. S. & H. Durrant-Whyte (1993). A decentralized Bayesian algorithm for identification of tracked targets. *IEEE Transactions on Systems, Man, and Cybernetics*, **23**, 1683-1698.

33. Pearl, J. (1986). On evidential reasoning in a hierarchy of hypotheses. *Artificial Intelligence*, **28**, 9-15.

34. Shafer, G. (1986). The combination of evidence. *International Journal of Intelligent Systems*, **1**, 155-179.

35. Cleckner, W. H. IV. (1985). *Tactical Evidential Reasoning: an Application of the Dempster-Shafer Theory of Evidence*. Master's Thesis, Naval Postgraduate School, Monterey, CA, 140 p.

36. Ng, K.-C. & B. Abramson (1990). Uncertainty management in expert systems. *IEEE Expert*, April 1990, 29-48.

37. Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, **38**, 325-339.

38. Cortes-Rello, E. & F. Golshani (1990). Uncertain reasoning using the Dempster-Shafer method: An application in forecasting and marketing management. *Expert Systems*, **7**, 9-17.

39. Yager, R. R. (1987). On the Dempster-Shafer framework and new combination rules. *Information Sciences*, **41**, 93-137.

40. Voorbraak, F. (1991). On the justification of Dempster's rule of combination. *Artificial Intelligence*, **48**, 171-197.

41. Yager, R. R. (1983). Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems*, **9**, 249-260.

42. Zadeh, L. A. (1984). Review of Shafer's "A mathematical theory of evidence". *AI Magazine*, Fall 1984, 81-83.

43. Zadeh, L. A. (1986). A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. *AI Magazine*, Summer 1986, 85-90.

44. Dubois, D. & H. Prade (1985). Combination and propagation of uncertainty with belief functions. *Proceedings 9th IJCAI*, Los Angeles, CA, August 18-23, 111-113.

45. Inagaki, T. (1991). Interdependence between safety-control policy and multiple-sensor schemes via Dempster-Shafer theory. *IEEE Transactions on Reliability*, **40**, 182-188.

46. Voorbraak, F. (1989). A computationally efficient approximation of Dempster-Shafer theory. *International Journal of Man-Machine Studies*, **30**, 525-536.

47. Barnett, J. A. (1981). Computational methods for a mathematical theory of evidence. *Proceedings of the Seventh IJCAI (International Joint Conference on AI)*, Morgan Kaufman, Palo Alto, CA, 868-875.

48. Gordon, J. & E. H. A. Shortliffe (1986). Method for managing evidential reasoning in a hierarchical hypothesis space. *Artificial Intelligence*, **26**, 323-357.

49. Shafer, G. (1985). Hierarchical evidence. *The Second Conference on Artificial Intelligence Applications*, 16-21.

50. Andress, K. M. & A. C. Kak (1988). Evidence accumulation and flow of control in a hierarchical spatial reasoning system. *AI Magazine*, **9**, 75-94.

51. Liu, L. J. & J. Y. Yang (1991). Model-based object classification using fused data. *SPIE Vol. 1611, Sensor Fusion IV*, 65-76.

52. Selzer, F. & D. Gutfinger (1988). LADAR and FLIR based sensor fusion for automatic target classification. *SPIE Vol. 1003, Sensor Fusion: Spatial Reasoning and Scene Interpretation*, 236-246.

53. Altoft, J. (1990). *RUNES: Reasoning in Uncertain Nested Evidence Spaces*. Master's Thesis. Carleton University, Ottawa, ON, 153 p.

54. Dasarathy, B. V. (1994). *Decision Fusion*. IEEE Computer Society Press. Los Alamitos, CA, 381 p.

55. Abdulghafour, M. & M. A. Abidi (1993). Data fusion through non-deterministic approaches - A comparison. *SPIE 2059, Sensor Fusion VI*, 37-53.

56. Jane's Fighting Ships 1984-85, Jane's Publishing Company Limited, England.

57. Jane's Weapons Systems 1983-84, Jane's Publishing Company Limited, England.

APPENDIX A
Examples of the Bayesian Approach

A.1 Examples of the Bayesian Approach (general case)

To better appreciate the Bayesian approach to uncertainty in terms of representation and combination of information, two simple applications are given.

The general context is the following: two possible missile types (type 1 and type 2) are known to be in the coverage area of two independent sensors. The first application deals with 2 successive identity declarations by a single sensor whereby the two declarations are missile type 1. Because these declarations are assumed to be conditionally independent, they can be fused using (3.6). If the likelihood matrix is given by

$$P(E_i|H_j) = \begin{pmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{pmatrix}$$

and if the *a priori* probabilities $P(H_j)$ are considered equal to 1/2, then the probability that each type of missile is present after the first declaration is:

$$P(\text{type }1 \mid E^1_{\text{type }1}) = \frac{0.8 \times 0.5}{0.8 \times 0.5 + 0.4 \times 0.5} = \frac{2}{3}$$
$$P(\text{type }2 \mid E^1_{\text{type }1}) = \frac{1}{3}$$

After the second declaration, we obtain the following *a posteriori* probabilities:

$$P(\text{type }1 \mid E^2_{\text{type }1}) = .8$$
$$P(\text{type }2 \mid E^2_{\text{type }1}) = .2$$

In the second application, two sensors (A and B) are capable of declaring the identity of targets and each sensor declares concurrently the missile to be of type 1. Let us assume that the sensors are identical such that the likelihood matrix of each sensor is given by:

$$P(E_i|H_j) = \begin{pmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{pmatrix}$$

Let us assume also that the *a priori* probabilities $P(H_j)$ are equal and that the sensors are independent, so that the joint probabilities are simply the product of their individual probabilities. The resulting likelihood matrix becomes:

$$\begin{pmatrix} 0.8 \times 0.8 & 0.4 \times 0.4 \\ 0.8 \times 0.2 & 0.4 \times 0.6 \\ 0.2 \times 0.8 & 0.6 \times 0.4 \\ 0.2 \times 0.2 & 0.6 \times 0.6 \end{pmatrix} = \begin{pmatrix} 0.64 & 0.16 \\ 0.16 & 0.24 \\ 0.16 & 0.24 \\ 0.04 & 0.36 \end{pmatrix}$$

where $0.64 =$ P(evidence from sensor A, evidence from sensor B | missile type 1)
$=$ P(evidence from sensor A | missile type 1) $\times$ P(evidence from sensor B | missile type 1)
$= 0.8 \times 0.8$
since evidences are conditionally independent.

The probability that each type of missile is present after both concurrent evidences is:

P(missile type 1 | Evidence sensor A, Evidence sensor B) =

$$\frac{0.64 \times 0.5}{0.64 \times 0.5 + 0.16 \times 0.5} = 0.8$$

P(missile type 2 | Evidence sensor A, Evidence sensor B) = 0.2

The fact that the same results are obtained in the two examples is not surprising since, in both cases, the successive evidences are assumed to be conditionally independent. Whether these evidences come from 2 different sources, or from the same source at two different points in time, makes no difference in the analysis.

A.2 Example of the Technique Suggested by J. Pearl

Figure 9 provides an example of a strict hierarchical tree of hypotheses: $\Omega = \{C, D, I, K, L, M, N, O, P\}$. As before, other letters are used to represent unions of these outcomes, e.g. $B = \{C, D\}$, $G = \{L, M\}$ and $H = \{K, N, O, P\}$. *A priori* probabilities are indicated for each set of interest. For example, $P(G) = P(L) + P(M) = 0.3 + 0.2 = 0.5$ and similarly $P(H) = 0.2$. Now suppose that information is received concerning hypothesis set H in such a way that

$$P(E_1 \mid H) = 0.5 \text{ and } P(E_1 \mid \overline{H}) = 0.2$$

This information can be represented differently by the likelihood matrix:

$$P(E_i|H_j) = \begin{pmatrix} 0.5 & 0.2 \\ 0.5 & 0.8 \end{pmatrix}$$

where $H_1 = H$, $H_2 = \overline{H}$ and $E_2 = \overline{E}$.
If E_1 is observed, then:

$$O(H) = 0.2 / 0.8 = 1/4$$
$$\lambda_H = 0.5 / 0.2 = 2.5$$
$$\alpha_H^t = 1 / [(2.5 \times 0.2) + 1 - 0.2] = 0.769$$
$$P(H \mid E_1) = 0.769 \times 2.5 \times 0.2 = 0.3845$$

and hence $P(H \mid E_1)$ represents the posterior probability of H given E_1, indicated in parentheses beside the prior probability of H, in Figure 9. To determine how this new evidence affects H's parents and children, we use Pearl's formulas. For J, which is a child of H, we find:

$$P(J \mid E_1) = 0.1 \times 0.769 \times 2.5 = 0.19225 \quad \text{(Child of H)}$$

For F, which is a parent of H, we find:

$$P(F \mid E_1) = 0.769\,(0.8 - 0.2) + 0.3845 = 0.8459 \quad \text{(Father of H).}$$

Updated values for every other parent and child of H are indicated in parentheses in Figure 9. The updated belief for hypothesis set A is not equal to one due to rounding-off errors.

It is important to note that Pearl's algorithm is based exclusively on the simple concept of proportional allocation. In the above example, for instance, node H is the only one that can be formally updated by Bayes' theorem, since the likelihood matrix merely specified $P(E_1|H)$ and $P(E_1|\overline{H})$. This information by itself does not permit updating of probabilities for the children or parents of H, even though it is known that these probabilities must also have changed. To alleviate this difficulty, Pearl suggests proportionally allocating $P(E_1|H)$ and $P(E_1|\overline{H})$ to the *a priori* evidence of the other nodes, while keeping in mind that each node of the tree should acquire a belief equal to the sum of the beliefs belonging to its immediate successors. Thus, for example, nodes J and K are both updated to 0.19225, on the basis that their *a priori* probabilities were equal to 0.1, despite the fact that no specific information is available to determine how these probabilities have actually been affected by evidence H.

While this rule is reasonable, it is clearly conventional and may not always lead to an appropriate estimation of the posterior probabilities associated with certain nodes of a hierarchy. On the other hand, adoption of such a rule is necessary in order that future evidence on node H, or on any other node, could again be incorporated into the probability distribution by Bayes' rule.

APPENDIX B
Examples of Evidential Theory

B.1 Example of Terminology of Evidential Theory

A simple numerical example will help clarify the wealth of terminology associated with the Evidential theory. Let $\Theta = \{X,Y,Z\}$. The set of all subsets of Θ (2^{Θ}) contains 8 elements, namely $\{X,Y,Z\},\{X,Y\},\{X,Z\},\{Y,Z\},\{X\},\{Y\},\{Z\},\varnothing$. Let us assign basic probability numbers to each subset as follows. This is formally the same as assigning probabilities to the preceding set of eight points, ignoring their nature, i.e., the fact that $\{X\} \subseteq \{X, Y, Z\}$, for example.

$m(\{X,Y,Z\}) = 0.1$	$m(\{X\}) = 0.2$
$m(\{X,Y\}) = 0.3$	$m(\{Y\}) = 0.0$
$m(\{X,Z\}) = 0.0$	$m(\{Z\}) = 0.1$
$m(\{Y,Z\}) = 0.3$	$m(\varnothing) = 0.0$

The focal elements are the following: $\{X,Y,Z\},\{X,Y\},\{Y,Z\},\{X\},\{Z\}$. They are the sets to which m assigns strictly positive mass. The degree of belief Bel for each subset is obtained as follows from (4.2):

$Bel(\{X,Y,Z\}) = 1.0$	$Bel(\{X\}) = 0.2$
$Bel(\{X,Y\}) = 0.5$	$Bel(\{Y\}) = 0.0$
$Bel(\{X,Z\}) = 0.3$	$Bel(\{Z\}) = 0.1$
$Bel(\{Y,Z\}) = 0.4$	$Bel(\varnothing) = 0.0$

Thus, for example, $Bel\{Y, Z\} = m\{Y, Z\} + m\{Y\} + m\{Z\} + m(\varnothing) = 0.4$.
Clearly, Bel adheres to the constraints of a belief function. As pointed out earlier, it is possible to retrieve basic probability numbers from the degrees of belief of each subset:
For example,

$$m(\{Y,Z\}) = (-1)^0\, Bel(\{Y,Z\}) + (-1)^1\, Bel(\{Y\}) + (-1)^1\, Bel(\{Z\}) + (-1)^2\, Bel(\{\varnothing\})$$
$$= 0.4 - 0.0 - 0.1 + 0.0 = 0.3$$

The commonality number Q for each subset is easily obtained from (4.4):

$Q(\{X,Y,Z\}) = 0.1$	$Q(\{X\}) = 0.6$
$Q(\{X,Y\}) = 0.4$	$Q(\{Y\}) = 0.7$
$Q(\{X,Z\}) = 0.1$	$Q(\{Z\}) = 0.5$
$Q(\{Y,Z\}) = 0.4$	$Q(\varnothing) = 1.0$

The same results can be obtained by applying (4.3). Thus, for instance:

$$Q(\{X,Y\}) = (-1)^2 \mathrm{Bel}(\overline{\{X,Y\}}) + (-1)^1 \mathrm{Bel}(\overline{\{X\}}) + (-1)^1 \mathrm{Bel}(\overline{\{Y\}}) + (-1)^0 \mathrm{Bel}(\overline{\varnothing})$$
$$= 0.1 - 0.4 - 0.3 + 1.0 = 0.4$$

Next, using (4.5), the degree of belief of each subset can be calculated from the commonality number. For example,

$$\mathrm{Bel}(\{Y,Z)\} = (-1)^1\, Q(\{X\}) + Q(\varnothing) = -0.6 + 1.0 = 0.4$$

To compute the degree of plausibility of each subset, (4.6) is used:

$\mathrm{Pl}(\{X,Y,Z\}) = 1.0$	$\mathrm{Pl}(\{X\}) = 0.6$
$\mathrm{Pl}(\{X,Y\}) = 0.9$	$\mathrm{Pl}(\{Y\}) = 0.7$
$\mathrm{Pl}(\{X,Z\}) = 1.0$	$\mathrm{Pl}(\{Z\}) = 0.5$
$\mathrm{Pl}(\{Y,Z\}) = 0.8$	$\mathrm{Pl}(\varnothing) = 0.0$

The evidential interval for each subset is then as follows:

subset $\{X,Y,Z\}$: [1.0, 1.0]	subset $\{X\}$: [0.2, 0.5]
subset $\{X,Y\}$: [0.5, 0.9]	subset $\{Y\}$: [0.0, 0.7]
subset $\{X,Z\}$: [0.3, 1.0]	subset $\{Z\}$: [0.1, 0.5]
subset $\{Y,Z\}$: [0.4, 0.8]	subset $\varnothing$: [0.0, 0.0]

B.2 Proof of eq. (4.12)

Proof: From the definition of A's commonality number (equation 4.4), one has

$$Q(A) = \sum_{\substack{B \subseteq \Theta \\ A \subseteq B}} m(B),$$

where $m = m_1 \oplus m_2$. Replacing m(B) with its orthogonal sum (equation 4.9) yields

$$Q(A) = K \sum_{\substack{B \subseteq \Theta \\ A \subseteq B}} \sum_{\substack{i,j \\ B_i \cap C_j = B}} m_1(B_i)\, m_2(C_j)$$

$$= K \sum_{\substack{i,j \\ A \subseteq B_i \cap C_j}} m_1(B_i)\, m_2(C_j)$$

$$= K \sum_{\substack{i,j \\ A \subseteq B_i \\ A \subseteq C_j}} m_1(B_i)\, m_2(C_j)$$

$$= K \left(\sum_{\substack{i \\ A \subseteq B_i}} m_1(B_i) \right) \left(\sum_{\substack{j \\ A \subseteq C_j}} m_2(C_j) \right)$$

$$= K \left(\sum_{\substack{B \subseteq \Theta \\ A \subseteq B}} m_1(B) \right) \left(\sum_{\substack{B \subseteq \Theta \\ A \subseteq B}} m_2(B) \right)$$

$$= K \ Q_1(A) \ Q_2(A), \quad \text{for all non-empty } A \subseteq \Theta.$$

This completes the proof.

B.3 Example of Dempster's Rule of Combination

Let $\Theta = \{X, Y, Z\}$, and m_1 and m_2 be basic probability assignments such that:

$m_1(\{X\}) = 0.2$	$m_2(\{Y\}) = 0.4$
$m_1(\{X, Y\}) = 0.4$	$m_2(\{X, Y\}) = 0.4$
$m_1(\Theta) = 0.4$	$m_2(\{X, Z\}) = 0.1$
	$m_2(\Theta) = 0.1$

and $m_i(A) = 0$, $i = 1,2$ for all non listed subsets of 2^{Θ}.

Using Dempster's rule of combination, we proceed as follows to derive a new basic probability assignment:

$$m = m_1 \oplus m_2$$

m_2	$m_1(\{X\})$ .2	$m_1(\{X,Y\})$ .4	$m_1(\Theta)$ .4
$m_2(\{Y\})$ .4	{} .08	{Y} .16	{Y} .16
$m_2(\{X,Y\})$ .4	{X} .08	{X,Y} .16	{X,Y} .16
$m_2(\{X,Z\})$ .1	{X} .02	{X} .04	{X,Z} .04
$m(\Theta)$ .1	{X} .02	{X,Y} .04	Θ .04

Before Normalization	After Normalization (1/K = 0.92)	
$m(\{X\}) = 0.16$	$m(\{X\}) = 0.17$	$Pl(\{X\}) = .65$
$m(\{Y\}) = 0.32$	$m(\{Y\}) = .35$	$Pl(\{Y\}) = 0.785$
$m(\{Z\}) = 0.0$	$m(\{Z\}) = 0.0$	$Pl(\{Z\}) = 0.09$
$m(\{X,Y\}) = 0.36$	$m(\{X,Y\}) = 0.39$	$Pl(\{X,Y\}) = 1.0$
$m(\{X,Z\}) = 0.04$	$m(\{X,Z\}) = 0.045$	$Pl(\{X,Z\}) = 0.65$
$m(\{Y,Z\}) = 0.0$	$m(\{Y,Z\}) = 0.0$	$Pl(\{Y,Z\}) = 0.83$
$m(\Theta) = 0.04$	$m(\Theta) = 0.045$	$Pl(\Theta) = 1.0$
$m(\varnothing) = 0.08$	$m(\varnothing) = 0.0$	$Pl(\varnothing) = 0.0$

Identical results in terms of the degree of plausibility are obtained from the commonality functions using (4.15). The commonality numbers calculated from (4.13) and the normalizing constant from (4.14) are given below:

$Q_1(\{X\}) = 1.0$	$Q_2(\{X\}) = 0.6$	$1/K = 0.92$
$Q_1(\{Y\}) = 0.8$	$Q_2(\{Y\}) = 0.9$	
$Q_1(\{Z\}) = 0.4$	$Q_2(\{Z\}) = 0.2$	
$Q_1(\{X,Y\}) = 0.8$	$Q_2(\{X,Y\}) = 0.5$	
$Q_1(\{X,Z\}) = 0.4$	$Q_2(\{X,Z\}) = 0.2$	
$Q_1(\{Y,Z\}) = 0.4$	$Q_2(\{Y,Z\}) = 0.1$	
$Q_1(\Theta) = 0.4$	$Q_2(\Theta) = 0.1$	
$Q_1(\varnothing) = 1.0$	$Q_2(\varnothing) = 1.0$	

B.4 Voorbraak's Example

In the example, the body of evidence Y is implied by X; as a consequence, evidence Y is already taken into account in the basic probability assignment m_X. The example is reproduced below.

> Let Θ be the frame of discernment $\{A, \overline{A}\}$, where A denotes the proposition "patient P has the flu". Suppose that X represents the observation that P has a fever > 39°C, that Y represents the observation that P has a fever > 38.5°C and that the basic probability assignments of X and Y are:
>
> $$m_X(A) = .6 \qquad m_X(\Theta) = .4$$
> $$m_Y(A) = .4 \qquad m_Y(\Theta) = .6$$
>
> According to Dempster's combination rule, $Bel_X(A) \oplus Bel_Y(A) = 0.76$. However, since Y is implied by X, we would assume that $Bel_X(A) \oplus Bel_Y(A) = Bel_X(A) = .6$. This is obviously not the case; therefore $Bel_X(A) \oplus Bel_Y(A)$ is unacceptable.

It is important to note that Voorbraak's example is based on the principle that an observation is received and inference is applied to the observation to obtain a conclusion:

$$\text{observation} \Rightarrow \text{inference} \Rightarrow \text{conclusion.}$$

Using a different terminology, we have:

$$\text{evidence} \Rightarrow \text{inference} \Rightarrow \text{hypothesis}.$$

This is clearly depicted by Voorbraak's example reproduced above:

$$\text{example}\begin{cases}\text{X: P has fever} > 39°\text{C} \Rightarrow \text{inference} \Rightarrow \text{A: patient P has flu}\\ \text{Y: P has fever} > 38.5°\text{C} \Rightarrow \text{inference} \Rightarrow \text{A: patient P has flu}\end{cases}$$

As mentioned earlier, $Bel_X(A) \oplus Bel_Y(A)$ is unacceptable in this context.

However, modifying Voorbraak's example by eliminating the inference process simplifies the independence concept, since the type of evidence has changed and the relationship between the evidence and inference process is eliminated. If we assume that doctors C and D provide independent diagnoses, then we could say that evidences X and Y are independent under this modified structure:

$$\text{modified example}\begin{cases}\text{X: doctor C diagnoses flu} \Rightarrow \text{A: patient P has flu}\\ \text{Y: doctor D diagnoses flu} \Rightarrow \text{A: patient P has flu}\end{cases}$$

Here $Bel_X(A) \oplus Bel_Y(A)$ is acceptable and the evidences seem independent according to Shafer since, when viewed abstractly, the information originates from two assumed independent diagnoses.

B.5 Example of Dempster's Rule with Too Coarse a Frame of Discernment

Let $\Theta = \{a, b, c, d\}$ and $\Omega = \{\omega_1, \omega_2\}$, where $\psi: 2^\Omega \rightarrow 2^\Theta$ is the *refining* given by $\psi(\{\omega_1\}) = \{a\}$ and $\psi(\{\omega_2\}) = \{b, c, d\}$. The set Ω is then a coarsening of Θ. Assume that the first body of evidence produced a simple support function S_1 over Θ focused on $A = \{a, b\}$, whereas the second body of evidence produced a simple support function S_2 over Θ focused on $B = \{a, c\}$. Neither S_1 nor S_2 record any support for either $\{a\}$ or $\{b, c, d\}$. Therefore, $S_1|2^\Omega$, $S_2|2^\Omega$ and $(S_1|2^\Omega) \oplus (S_2|2^\Omega)$ are all vacuous. However, $((S_1|2^\Theta) \oplus (S_2|2^\Theta))|2^\Omega$ is not vacuous, since it provides the degree of support $S_1(A)\, S_2(B) > 0$ for $A \cap B = \{a\}$. Thus $((S_1|2^\Theta) \oplus (S_2|2^\Theta))|2^\Omega \neq ((S_1|2^\Omega) \oplus (S_2|2^\Omega))|2^\Omega$.

In our study of identity declaration fusion, this difficulty can be easily alleviated by choosing the frame of discernment fine enough to discern all relevant interaction of evidence to be combined. This will be easily attainable due to the nature of the evidence (identity declaration) and because the evidence will be structured in a strict hierarchy.

B.6 Proof of Equation (4.16)

Proof:

a. For any A, one has $n_A \leq n$, and hence

$$P_m \geq \frac{1}{n}\sum_A m(A).$$

Now since $\sum m(A) = 1$, it follows that $P_m \geq 1/n$. Also, $n_A \geq 1$ for any $A \neq \varnothing$, and hence

$$P_m \leq \sum_{A \subset X} m(A) \leq 1.$$

b. If m is vacuous, m(X)=1 and hence $P_m = 1/n$.
If m is not vacuous, then there exists some A such that $m(A) > 0$ and $n_A < n$; therefore

$$P_m \geq \frac{1}{n}.$$

c. Assume that m is Bayesian. Then the sets having $m(A) > 0$ are only the singletons. Thus

$$P_m = \sum_{i=1}^{n} m(\{x_i\}) \leq 1.$$

Assume that m is not Bayesian. Then there exists some A such that $m(A) > 0$ and $n_A > 1$, whence $P_m < 1$.
This completes the proof.

As an example, let $\Theta = \{W, X, Y, Z\}$, and m_1 and m_2 be basic probability assignments defined as follows:

$m_1(\{X,Y\}) = 0.4$ $\quad$ $m_2(\{W, X, Y\}) = 0.4$
$m_1(\{Z\}) = 0.2$ $\quad$ $m_2(\{Z\}) = 0.2$
$m_1(\Theta) = 0.4$ $\quad$ $m_2(\Theta) = 0.4$

Again, $m_i(A) = 0$, $i = 1,2$ if A is not listed above.
Therefore,

$$P_{m_1} = 0.4/2 + 0.2/1 = 0.4 \qquad P_{m_2} = 0.4/3 + 0.2/1 = 0.33$$

B.7 Proofs of the Entropy Measure Properties

Proof:

a. Since for a Bayesian belief function m(A) = 0 for all non-singletons,

$$E_m = \sum_{x \subset X} m(\{x\}) \, \mathrm{Con}(\mathrm{Bel}, \mathrm{Bel}_A).$$

Let g_x denote the basic assignment function associated with the certain support function at $\{x\}$. Then

$g_x(\{x\}) = 1,$
$g_x(B) = 0$ for all other $B \subset X$, and
$\mathrm{Con}(\mathrm{Bel}, \mathrm{Bel}_{\{x\}}) = -\ln(1-k)$ where $k = \sum_{\substack{i,j \\ \text{for } A_i \cap B_j = \varnothing}} m(A_i)\, g_x(B_j).$

Since $g_x(B) = 0$ for $B \neq \{x\}$ and is equal to 1 elsewhere,

$$k = \sum_{\substack{i \\ \text{for } A_i \cap \{x\} = \varnothing}} m(A_i).$$

Since m is Bayesian,

$$k = \sum_{\substack{i \\ \{x_i\} \cap \{x\} = \varnothing}} m(\{x_i\}) = \sum_{\substack{i \\ \text{for } x_i \neq x}} m(\{x_i\}) = 1 - m\{(x\}).$$

Thus,

$$\mathrm{Con}(\mathrm{Bel}, \mathrm{Bel}_{\{x\}}) = -\ln(1 - [1 - m(\{x\})]) = -\ln(m[\{x\})],$$

and hence

$$E_m = -\sum_{x \in X} m(\{x\}) \ln[m(\{x\})].$$

b. Since $\mathrm{Pl}(A) \in [0,1]$ for all $A \subset X$, one has $\ln(\mathrm{Pl}(A)) \leq 0$. Furthermore, since $m(A) \in [0,1]$, it must be that

$$E_m = -\sum_{A \subset X} m(A) \ln(Pl(A)) \geq 0.$$

Let us introduce n focal elements with values $m(A_i) = a_i$. We have

$$E_m = -\sum_{i=1}^{n} m(A_i) \ln[Pl(A_i)]$$

where

$$Pl(A_i) = \sum_{\substack{j \\ \text{for } A_i \cap A_j \neq \varnothing}} m(A_j) + \sum_{\substack{A_j \\ \text{for } A_i \cap A_j \neq \varnothing \\ i=j}} m(A_j)$$

$$= m(A_i) + d_i = a_i + d_i.$$

Therefore, one has

$$E_m = -\sum_{i=1}^{n} a_i \ln(a_i + d_i).$$

As d_i increases, $\ln(a_i + d_i)$ increases and $-\sum_{i=1}^{n} \ln(a_i + d_i)$ decreases. Consequently, E_m is maximal when $d_i = 0$ for all i. This occurs when all the A_i are disjoint. If we assume n disjoint focal elements with $m(A_i) = 1/n$, we obtain a maximal E_m, namely

$$E_m = -\sum_{i=1}^{n} 1/n \ln(1/n) = \ln(n).$$

c. From the definition of E_m, $E_m = 0$ if there is an A such that $m(A) \neq 0$, which requires in turn that $\ln[Pl(A)] = 0$ and Pl(A)=1. Since

$$Pl(A) = \sum_{\substack{B \\ B \cap A \neq \varnothing}} m(B),$$

this means that every pair of focal elements must have at least one element in common.

d. $E_m = \ln(n)$ iff $m(A_i) = 1/n$ for i=1, 2,...,n was proved in b. above.

This completes the proof.

Shannon's entropy measures the discordance associated with a probability distribution (Yager, 1983). As an example, let $\Theta = \{W, X, Y, Z\}$, and m_1 and m_2 be basic probability assignments defined as follows:

$$m_1(\{W\}) = 0.25 \qquad m_2(\{W\}) = 0.5$$
$$m_1(\{X\}) = 0.25 \qquad m_2(\{W, X\}) = 0.25$$
$$m_1(\{Y\}) = 0.25 \qquad m_2(\{Z\}) = 0.25$$
$$m_1(\{Z\}) = 0.25$$

Then,

$$E_{m_1} = -[0.25 \bullet \ln(0.25) + 0.25 \bullet \ln(0.25) + 0.25 \bullet \ln(0.25) + 0.25 \bullet \ln 0(.25)] = 1.386$$
$$E_{m_2} = -[0.5 \bullet \ln(0.75) + 0.25 \bullet \ln(0.75) + 0.25 \bullet \ln(0.25)] = 0.562$$

$$E_{m_2} = -[0.5 \bullet \ln(0.75) + 0.25 \bullet \ln(0.75) + 0.25 \bullet \ln(0.25)] = 0.562$$

APPENDIX C
The Shafer and Logan Algorithm

C.1 Formulas for the Shafer and Logan Algorithm

The formulas for the Shafer and Logan algorithm are given below. For each node A in $\mathfrak{R}$, let

$$A_0^+ = \mathrm{Bel}_A(A) \qquad A_0^- = \mathrm{Bel}_A(\overline{A})$$
$$A_\downarrow^+ = \mathrm{Bel}_A^\downarrow(A) \qquad A_\downarrow^- = \mathrm{Bel}_A^\downarrow(\overline{A})$$
$$A^+ = (\mathrm{Bel}_A \oplus \mathrm{Bel}_A^\downarrow)(A) \qquad A^- = (\mathrm{Bel}_A \oplus \mathrm{Bel}_A^\downarrow)(\overline{A})$$
$$A_\Diamond^+ = (\mathrm{Bel}_A \oplus \mathrm{Bel}_A^\Diamond)(A) \qquad A_\Diamond^- = (\mathrm{Bel}_A \oplus \mathrm{Bel}_A^\Diamond)(\overline{A})$$
$$A_\ominus^+ = \mathrm{Bel}_\ominus^\downarrow(A) \qquad A_\ominus^- = \mathrm{Bel}_\ominus^\downarrow(\overline{A})$$

Stage 1

Calculate $A_\downarrow^+$ and $A_\downarrow^-$ from B^+ and B^- for B in ℓ_A:

$$A_\downarrow^+ = 1\text{-}K,$$
$$A_\downarrow^- = K \bullet \prod_{B\in\ell_A} B^- / (1-B^+)$$
where $1/K = 1 + \sum_{B\in\ell_A} B^+ / (1-B^+)$.

Calculate A^+ and A^- from A_0^+, A_0^-, $A_\downarrow^+$ and $A_\downarrow^-$:

$$A^+ = 1 - K(1-A_0^+)(1-A_\downarrow^+),$$
$$A^- = 1 - K(1-A_0^-)(1-A_\downarrow^-),$$
where $1/K = 1 - A_0^+ \bullet A_\downarrow^- - A_0^- \bullet A_\downarrow^+$.

Stage 2

Calculate $A_\ominus^+$ and $A_\ominus^-$ for A in $\ell_\ominus$ from A^+ and A^- for A in $\ell_\ominus$:

$$A_\ominus^+ = 1 - K(1 + \sum_{\substack{B\in\ell_\ominus \\ B\neq A}} B^+ / (1-B^+)) - \prod_{\substack{B\in\ell_\ominus \\ B\neq A}} B^- / (1-B^+),$$

$$A_\ominus^- = 1 - K(1-A^-) / (1-A^+),$$

where $1/K = 1 + \sum_{B \in \ell_\Theta} B^+/(1-B^+) - \prod_{B \in \ell_\Theta} B^-/(1-B^+)$.

Stage 3

Calculate $A_\Diamond^+$ and $A_\Diamond^-$ from A_Θ^+, A_Θ^-, $A_\downarrow^+$, and $A_\downarrow^-$:

$$A_\Diamond^+ = 1 - K(1 - A_\Theta^+)/(1 - A_\downarrow^+),$$

$$A_\Diamond^- = 1 - K(1 - A_\Theta^-)/(1 - A_\downarrow^-),$$

where $1/K = \dfrac{1 - A_\Theta^+}{1 - A_\downarrow^+} + \dfrac{1 - A_\Theta^-}{1 - A_\downarrow^-} - \dfrac{1 - A_\Theta^+ - A_\Theta^-}{1 - A_\downarrow^+ - A_\downarrow^-}$.

Calculate B_A^+, B_A^-, and B_A^* from C^+ and C^- for C in ℓ_A:

$$B_A^+ = 1 - K(1 + \sum_{\substack{C \in \ell_A \\ C \neq B}} C^+/(1 - C^+)),$$

$$B_A^- = 1 - K(1 - B^-)/(1 - B^+),$$

$$B_A^* = 1 - K(1 + \sum_{\substack{C \in \ell_A \\ C \neq B}} C^+/(1 - C^+) - \prod_{\substack{C \in \ell_A \\ C \neq B}} C^-/(1 - C^+)),$$

where $1/K = 1 + \sum_{C \in \ell_A} C^+/(1 - C^+)$.

Calculate

B_Θ^+ and B_Θ^- from $A_\downarrow^+$, $A_\downarrow^-$, $A_\Diamond^+$, $A_\Diamond^-$, B_A^+, B_A^- and B_A^* where B is a daughter of A:

$$B_\Theta^+ = K(A_\Diamond^+(B_A^* - A_\downarrow^-) + (1 - A_\Diamond^+ - A_\Diamond^-)B_A^+),$$

$$B_\Theta^- = 1 - K(1 - A_\Diamond^-)(1 - B_A^-),$$

where $1/K = 1 - A_\downarrow^+ \cdot A_\Diamond^- - A_\downarrow^- \cdot A_\Diamond^+$.

C.2 Example of Dempster's Combination Rule with Bayesian Belief Functions

This example uses the strict hierarchical tree illustrated in Figure 9 of Subsection 3.6.3. The *a priori* probabilities are indicated for each set of interest. Information received concerns hypothesis H:

$$P(E_1|H) = 0.5 \text{ and } P(E_1|\overline{H}) = 0.2$$

To be compatible with the input of Dempster's combination rule, this information must be transformed as follows:

$$P(H|E_1) = \frac{P(E_1|H)}{P(E_1|H) + P(E_1|\overline{H})}$$

such that $P(H|E_1) = 0.714$ and $P(\overline{H}|E_1) = 0.286$. We therefore obtain the following basic probability assignments:

$$m_1(\{H\}) = (\{K, N, O, P\}) = 0.714,$$
$$m_1(\{\overline{H}\}) = (\{C, D, I, L, M\}) = 0.286,$$
$$m_1(\Theta) = (\{C, D, I, K, L, M, N, O, P\}) = 0.0,$$

which represent a simple support function focused on a subset of Θ and its complement, with no uncertainty. This basic probability assignment is then combined to the *a priori* probability of hypothesis H (0.2) using Dempster's combination rule: $m = m_1 \oplus m_2$:

$$m_1(\{H\}) = m_1(\{K, N, O, P\}) = 0.7143$$
$$m_1(\{\overline{H}\}) = m_1(\{C, D, I, L, M\}) = 0.2857$$
$$m_1(\Theta) = 0.0$$
$$m_2(\{H\}) = m_2(\{K, N, O, P\}) = 0.2$$
$$m_2(\{G\}) = m_2(\{L, M\}) = 0.5$$
$$m_2(\{I\}) = 0.1$$
$$m_2(\{B\}) = m_2(\{C, D\}) = 0.2$$

Before Normalization	After Normalization ($K^{-1} = 0.37142$)
$m(\{K, N, O, P\}) = 0.14286$	$m(\{K, N, O, P\}) = 0.3845$
$m(\{L, M\}) = 0.14285$	$m(\{L, M\}) = 0.3845$
$m(\{I\}) = 0.02857$	$m(\{I\}) = 0.0769$
$m(\{C, D\}) = 0.05714$	$m(\{C, D\}) = 0.1538$
$m(\varnothing) = 0.62858$	$m(\varnothing) = 0.0$

C.3 Example of Shafer & Logan Algorithm

This example, in which 6 sets of evidence are combined, illustrates the propagation effect of the Shafer & Logan algorithm. The same strict hierarchical tree as above is used. The example is composed of 6 steps, at which additional evidence is received for a specific node in the form of a simple support function or dichotomous function, and then combined. The new belief (Bel) and plausibility (Pl) values are calculated for each node. As a reminder, for each node A in the tree:

$$\begin{aligned}
\mathrm{Bel}(A) &= \mathrm{Bel}_{\Theta}^{\downarrow}(A) \\
&= A_{\Theta}^{+} \text{ (according to Annex B)}, \\
\mathrm{Pl}(A) &= 1-\mathrm{Bel}(\overline{A}) = 1-\mathrm{Bel}_{\Theta}^{\downarrow}(\overline{A}) \\
&= 1-A_{\Theta}^{-} \text{ (according to Annex B)}.
\end{aligned}$$

The evidence to be combined is as follows:

$$\begin{array}{lll}
\text{step1: } m(\{N\}) = .5 & m(\Theta) = .5 & \\
\text{step2: } m(\{H\}) = .5 & m(\Theta) = .5 & \\
\text{step3: } m(\{F\}) = .9 & m(\Theta) = .1 & \\
\text{step4: } m(\{F\}) = .0 & m(\{\overline{F}\}) = .5 & m(\Theta) = .5 \\
\text{step5: } m(\{B\}) = .99 & m(\{F\}) = .01 & m(\Theta) = .0 \\
\text{step6: } m(\{B\}) = .05 & m(\Theta) = .95 &
\end{array}$$

At step 0, all the belief and plausibility values of each node are zero. Figures 13 to 18 show the results of the Shafer-Logan algorithm after adding evidence from step 1 to 6 respectively.

To demonstrate the use of the various formulas of the Shafer-Logan algorithm given in Section C.1, calculations are shown below for step 1.

Stage 0

$m(\{N\}) = \mathrm{Bel}(\{N\}) = .5$ and $m(\Theta) = .5$

Stage 1

Let A = node J; then

$1/K = 1 + .5/(1-.5) + 0/1 + 0/1 = 2; K = .5$

$J_{\downarrow}^{+} = 1 - K = .5$

$J_{\downarrow}^{-} = .5\,(0/(1-.5) \times 0/1 \times 0/1) = 0$

$1/K = 1 - 0 \times 0 - 0 \times .5 = 1; K = 1$

$J^{+} = 1 - 1\,(1-0)(1-.5) = .5$

$J^{-} = 1 - 1\,(1-0)(1-0) = 0$

Let A = node H; then

$1/K = 1 + .5/(1-.5) + 0/1 = 2; K = .5$

$H_{\downarrow}^{+} = 1 - K = .5$

$H_{\downarrow}^{-} = .5\,(0/(1-.5) \times 0/1) = 0$

$1/K = 1 - 0 \times 0 - 0 \times .5 = 1; K = 1$

$H^{+} = 1 - 1\,(1-0)(1-.5) = .5$

$H^{-} = 1 - 1\,(1-0)(1-0) = 0$

Let A = node F; then

$1/K = 1 + 0/1 + .5/(1-.5) + 0/1 = 2; K = .5$

$F_{\downarrow}^{+} = 1 - K = .5$

$F_{\downarrow}^{-} = .5\,(0/1 \times 0/(1-.5) \times 0/1) = 0$

$1/K = 1 - 0 \times 0 - 0 \times .5 = 1; K = 1$

$F^{+} = 1 - 1\,(1-0)(1-.5) = .5$

$F^{-} = 1 - 1\,(1-0)(1-0) = 0$

Let A = node G; then

$G_{\downarrow}^{+} = G_{\downarrow}^{-} = G^{+} = G^{-} = 0$

Let A = node B; then

$B_{\downarrow}^{+} = B_{\downarrow}^{-} = B^{+} = B^{-} = 0$

Stage 2

$1/K = 1 + (0/1 + .5/(1-.5)) - (0/1 \times 0/.5) = 2; K = .5$
$B_{\Theta}^{+} = 1 - .5\,(1 + .5/(1-.5)) - 0 = 0$
$B_{\Theta}^{-} = 1 - .5\,(1-0)/(1 - 0) = .5$
$F_{\Theta}^{+} = 1 - .5\,(1 + 0) - 0 = .5$
$F_{\Theta}^{-} = 1 - .5\,(1 - 0)/(1 - .5) = 0$

Stage 3

Let A = node B; then
$1/K = (1 - 0)/(1 - 0) + (1 - .5)/(1 - 0) - (1- 0 - .5)/(1 - 0 - 0) = 1$
$B_{\Diamond}^{+} = 1 - 1(1 - 0)/(1 - 0) = 0$
$B_{\Diamond}^{-} = 1 - 1(1 - .5)/(1 - 0) = .5$

$1/K = 1 + 0/1 + 0/1 = 1; K = 1$
$C_{B}^{+} = 1 - 1(1 + 0/(1 - 0)) = 0$
$C_{B}^{-} = 1 - 1(1 - 0)/(1 - 0) = 0$
$C_{B}^{*} = 1 - 1(1 + 0/1 - 0/1) = 0$
and in a similar fashion,
$D_{B}^{+} = D_{B}^{-} = D_{B}^{*} = 0$

$1/K = 1 + 0 \times .5 - 0 \times 0 = 1$
$C_{\Theta}^{+} = 1\,(0 \times (0 - 0) + (1 - 0 - .5) \times 0 = 0$
$C_{\Theta}^{-} = 1 - 1(1 - .5)(1 - 0) = .5$
and in a similar fashion,
$D_{\Theta}^{+} = 0$ and $D_{\Theta}^{-} = .5$

Let A = node F; then

$1/K = (1 - .5)/(1 - .5) + (1 - 0)/(1 - 0) - (1- .5 - 0)/(1 - .5 - 0) = 1$
$F_{\Diamond}^{+} = 1 - 1(1 - .5)/(1 - .5) = 0$
$F_{\Diamond}^{-} = 1 - 1(1 - 0)/(1 - 0) = 0$
$1/K = 1 + 0/(1-0) + .5/(1-.5) + 0/(1-0) = 2; K = .5$

$G_F^{+} = 1 - .5(1 + .5/(1 - .5) + 0/(1 - 0)) = 0$
$G_F^{-} = 1 - .5(1 - 0)/(1 - 0) = .5$
$G_F^{*} = 1 - .5(1 + .5/(1 - .5) + 0/(1 - 0) -0) = 0$
$H_F^{+} = 1 - .5(1 + 0/(1 - 0) + 0/(1 - 0)) = .5$
$H_F^{-} = 1 - .5(1 - 0)/(1 - .5) = 0$
$H_F^{*} = 1 - .5(1 + 0/(1 - 0)+ 0/(1 - 0) -0) = .5$
$I_F^{+} = 1 - .5(1 + 0/(1 - 0)+ .5/(1 - .5)) = 0$
$I_F^{-} = 1 - .5(1 - 0)/(1 - 0) = .5$
$I_F^{*} = 1 - .5(1 + 0/(1 - 0)+ .5/(1 - .5) -0) = 0$

$1/K = 1 + .5 \times 0 - 0 \times 0 = 1$
$G_{\Theta}^{+} = 1 (0 \times (0 - 0) + (1 - 0 - 0) \times 0) = 0$
$G_{\Theta}^{-} = 1 - 1(1 - 0)(1 - .5) = .5$
$H_{\Theta}^{+} = 1 (0 \times (.5 - 0) + (1 - 0 - 0) \times .5) = .5$
$H_{\Theta}^{-} = 1 - 1(1 - 0)(1 - 0) = 0$
$I_{\Theta}^{+} = 1 (0 \times (0 - 0) + (1 - 0 - 0) \times 0) = 0$
$I_{\Theta}^{-} = 1 - 1(1 - 0)(1 - .5) = .5$

We have thus obtained Bel(B), Bel(C), Bel(D), Bel(F), Bel(G), Bel(H), Bel(I), where

$$\mathrm{Bel(B)} = \mathrm{Bel}_{\Theta}^{\downarrow}(B) = B_{\Theta}^{+} = 0 \text{ and}$$
$$\mathrm{Pl(B)} = 1 - \mathrm{Bel}_{\Theta}^{\downarrow}(\overline{B}) = 1 - B_{\Theta}^{-} = 1 - .5 = .5 .$$

The belief and plausibility of the other nodes can be obtained in a similar fashion using stage 3 of the algorithm.

INTERNAL DISTRIBUTION

DREV - R - 9527

1 - Deputy Director General
1 - Chief Scientist
6 - Document Library
1 - É. Bossé (author)
1 - J.M.J. Roy (author)
1 - R. Carling
1 - B. Chalmers
1 - G. Picard
1 - G. Otis
1 - LCdr S. Dubois
1 - J.-C. Labbé
1 - P. Labbé
1 - M. Gauvin
1 - D. Demers
1 - S. Paradis
1 - G. Thibault
1 - J.M. Thériault

UNCLASSIFIED
SECURITY CLASSIFICATION OF FORM
(Highest classification of Title, Abstract, Keywords)

DOCUMENT CONTROL DATA

1. ORIGINATOR (name and address) DREV 2459 Blvd Pie XI North Val Bélair, Qc G3J 1X5 CANADA	2. SECURITY CLASSIFICATION (Including special warning terms if applicable) UNCLASSIFIED

3. TITLE (Its classification should be indicated by the appropriate abbreviation (S,C,R or U)

FUSION OF HIERARCHICAL IDENTITY DECLARATIONS FOR NAVAL COMMAND AND CONTROL (U)

4. AUTHORS (Last name, first name, middle initial. If military, show rank, e.g. Doe, Maj. John E.)

DES GROSEILLIERS, Lyne, BOSSÉ, Éloi and ROY, Jean

5. DATE OF PUBLICATION (month and year)	6a. NO. OF PAGES	6b. NO. OF REFERENCES
1996	110	57

7. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. Give the inclusive dates when a specific reporting period is covered.)

REPORT

8. SPONSORING ACTIVITY (name and address)

DREV

9a. PROJECT OR GRANT NO. (Please specify whether project or grant)	9b. CONTRACT NO.
1ae12 (Sensor Data Fusion)	N/A

10a. ORIGINATOR'S DOCUMENT NUMBER	10b. OTHER DOCUMENT NOS.
DREV R-9527	N/A

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

- [X] Unlimited distribution
- [ ] Contractors in appoved countries (specify)
- [ ] Canadian contractors (with need-to-know)
- [ ] Government (with need-to-know)
- [ ] Defence departments
- [ ] Other (please specify) :

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

UNCLASSIFIED
SECURITY CLASSIFICATION OF FORM

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

Within the context of naval warfare, commanders and their staff require access to a wide range of information to carry out their duties. This information provides them with the knowledge necessary to determine the position, identity and behavior of the enemy. This document is concerned with the fusion of identity declarations through the use of statistical analysis rooted in the Dempster-Shafer theory of evidence. It proposes to hierarchically structure the declarations according to STANAG 4420 (Display Symbology and Colours for NATO Maritime Units). More specifically the aim of this document is twofold: to explore the problem of fusing identity declarations emanating from different sources, and to offer the decision maker a quantitative analysis based on statistical methodology that can enhance his/her decision making process regarding the identity of detected objects.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it is not possible to select indexing terms which are Unclassified, the classification of each sould be indicated as with the title.)

Information Fusion
Identification
Bayesian Theory
Evidential Theory
Dempster-Shafer Theory

Requests for documents
should be sent to:

DIRECTOR SCIENTIFIC INFORMATION SERVICES
Dept. of National Defence
Ottawa, Ontario
K1A 0K2
Tel: (613) 995-2971
Fax: (613) 996-0392

NO. OF COPIES NOMBRE DE COPIES	COPY NO. COPIE N°	INFORMATION SCIENTIST'S INITIALS INITIALES DE L'AGENT D'INFORMATION SCIENTIFIQUE
1	1	BA
AQUISITION ROUTE FOURNI PAR ▶		DREV
DATE ▶		10 Oct 96
DSIS ACCESSION NO. NUMÉRO DSIS ▶		

DND 1158 (6-87)

National Defence Défense nationale

499630

PLEASE RETURN THIS DOCUMENT TO THE FOLLOWING ADDRESS:

DIRECTOR
SCIENTIFIC INFORMATION SERVICES
NATIONAL DEFENCE
HEADQUARTERS
OTTAWA, ONT. - CANADA K1A 0K2

PRIÈRE DE RETOURNER CE DOCUMENT À L'ADRESSE SUIVANTE:

DIRECTEUR
SERVICES D'INFORMATION SCIENTIFIQUES
QUARTIER GÉNÉRAL
DE LA DÉFENSE NATIONALE
OTTAWA, ONT. - CANADA K1A 0K2

Toute demande de document
doit être adressée à:

DIRECTEUR - SERVICES D'INFORMATION SCIENTIFIQUE
Ministère de la Défense nationale
Ottawa, Ontario
K1A 0K2

Téléphone: (613) 995-2971
Télécopieur: (613) 996-0392